



## Multi-proxy analyses of a mid-15th century Middle Iron Age Bantu-speaker palaeo-faecal specimen elucidates the configuration of the 'ancestral' sub-Saharan African intestinal microbiome

Rifkin, Riaan F.; Vikram, Surendra; Ramond, Jean Baptiste; Rey-Iglesia, Alba; Brand, Tina B.; Porraz, Guillaume; Val, Aurore; Hall, Grant; Woodborne, Stephan; Le Bailly, Matthieu; Potgieter, Marnie; Underdown, Simon J.; Koopman, Jessica E.; Cowan, Don A.; Van De Peer, Yves; Willerslev, Eske; Hansen, Anders J.

*Published in:*  
Microbiome

*DOI:*  
[10.1186/s40168-020-00832-x](https://doi.org/10.1186/s40168-020-00832-x)

*Publication date:*  
2020

*Document version*  
Publisher's PDF, also known as Version of record

*Document license:*  
[CC BY](#)

*Citation for published version (APA):*  
Rifkin, R. F., Vikram, S., Ramond, J. B., Rey-Iglesia, A., Brand, T. B., Porraz, G., Val, A., Hall, G., Woodborne, S., Le Bailly, M., Potgieter, M., Underdown, S. J., Koopman, J. E., Cowan, D. A., Van De Peer, Y., Willerslev, E., & Hansen, A. J. (2020). Multi-proxy analyses of a mid-15th century Middle Iron Age Bantu-speaker palaeo-faecal specimen elucidates the configuration of the 'ancestral' sub-Saharan African intestinal microbiome. *Microbiome*, 8(1), [62]. <https://doi.org/10.1186/s40168-020-00832-x>

RESEARCH

Open Access



# Multi-proxy analyses of a mid-15th century Middle Iron Age Bantu-speaker palaeo-faecal specimen elucidates the configuration of the 'ancestral' sub-Saharan African intestinal microbiome

Riaan F. Rifkin<sup>1,2\*</sup>, Surendra Vikram<sup>1†</sup>, Jean-Baptiste Ramond<sup>1,2,3</sup>, Alba Rey-Iglesia<sup>4</sup>, Tina B. Brand<sup>4</sup>, Guillaume Porraz<sup>5,6</sup>, Aurore Val<sup>6,7</sup>, Grant Hall<sup>8</sup>, Stephan Woodborne<sup>8,9</sup>, Matthieu Le Bailly<sup>10</sup>, Marnie Potgieter<sup>1</sup>, Simon J. Underdown<sup>1,2</sup>, Jessica E. Koopman<sup>1</sup>, Don A. Cowan<sup>1</sup>, Yves Van de Peer<sup>1,11,12</sup>, Eske Willerslev<sup>4,13,14</sup> and Anders J. Hansen<sup>4\*</sup>

## Abstract

**Background:** The archaeological incidence of ancient human faecal material provides a rare opportunity to explore the taxonomic composition and metabolic capacity of the ancestral human intestinal microbiome (IM). Here, we report the results of the shotgun metagenomic analyses of an ancient South African palaeo-faecal specimen.

**Methods:** Following the recovery of a single desiccated palaeo-faecal specimen from Bushman Rock Shelter in Limpopo Province, South Africa, we applied a multi-proxy analytical protocol to the sample. The extraction of ancient DNA from the specimen and its subsequent shotgun metagenomic sequencing facilitated the taxonomic and metabolic characterisation of this ancient human IM.

**Results:** Our results indicate that the distal IM of the Neolithic 'Middle Iron Age' (c. AD 1460) Bantu-speaking individual exhibits features indicative of a largely mixed forager-agro-pastoralist diet. Subsequent comparison with the IMs of the Tyrolean Iceman (Ötzi) and contemporary Hadza hunter-gatherers, Malawian agro-pastoralists and Italians reveals that this IM precedes recent adaptation to 'Western' diets, including the consumption of coffee, tea, chocolate, citrus and soy, and the use of antibiotics, analgesics and also exposure to various toxic environmental pollutants.

(Continued on next page)

\* Correspondence: [riaanrifkin@gmail.com](mailto:riaanrifkin@gmail.com); [ajhansen@snm.ku.dk](mailto:ajhansen@snm.ku.dk)

<sup>†</sup>Riaan F. Rifkin and Surendra Vikram contributed equally to this work.

<sup>1</sup>Center for Microbial Ecology and Genomics, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Hatfield, South Africa

<sup>4</sup>Centre for GeoGenetics, GLOBE Institute, University of Copenhagen, Hatfield, Denmark

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Conclusions:** Our analyses reveal some of the causes and means by which current human IMs are likely to have responded to recent dietary changes, prescription medications and environmental pollutants, providing rare insight into human IM evolution following the advent of the Neolithic c. 12,000 years ago.

**Keywords:** Ancient DNA, Human evolution, Molecular ecology, Intestinal microbiome, Taxonomic composition, Metabolic capacity

## Background

The human gastrointestinal tract (GI) harbours a dynamic population of bacteria, archaea, fungi, protozoa and viruses; the intestinal microbiota. This collection of microorganisms, comprising the human intestinal microbiome (IM) [1], performs critical functions in digestion, development, behaviour and immunity [2, 3]. Modifications of the core IM composition (dysbiosis) have been associated with the pathogenesis of inflammatory diseases and infections [3, 4], including autoimmune and allergic diseases, obesity, inflammatory bowel disease and diabetes [5]. Despite its clinical importance, the factors that contribute to changes in IM taxonomic composition and functionality are not entirely understood [6, 7]. This is attributed to the fact that most of what is known about the human IM is based on contemporary industrialised and ‘traditional’ human societies [8–10]. In evolutionary terms, our species have subsisted by hunting and gathering for > 90% of our existence [11]. Evidence derived from the analyses of the IMs of traditional societies, including the Tanzanian Hadza hunter-gatherers [8], the Venezuelan Yanomami Amerindians [5], the BaAka Pygmies in the Central African Republic [12] and the Arctic Inuit [13], is thus widely viewed as representing ‘snapshots’ of ancient human IM composition. However, as exposure to Western diets, medicines and microbes cannot be excluded, one must be cautious about the use of these ethnographic cohorts as proxies for prehistoric human IMs [14].

The transformation of the IMs of present-day humans to their current ‘modernised’ state commenced millennia ago, with the advent of the Neolithic, which, at c. 12,000 years ago (ya), resulted in the first major human dietary transition [15]. But precisely how our IMs changed following the advent of the Neolithic, and the Industrial Revolution after c. AD 1760, remains ambiguous [16–18]. In this regard, the analyses of ancient human IMs provide a unique view into the co-evolution of microbes and human hosts, host microbial ecology and changing human IM-related health states through time [2]. Indeed, over the past 15 million years, multiple lineages of intestinal bacterial taxa arose via co-speciation with African hominins and non-human primates, i.e. chimpanzees, bonobos and gorillas [19]. The departure of behaviourally ‘fully-modern’ *Homo sapiens* from Africa c.

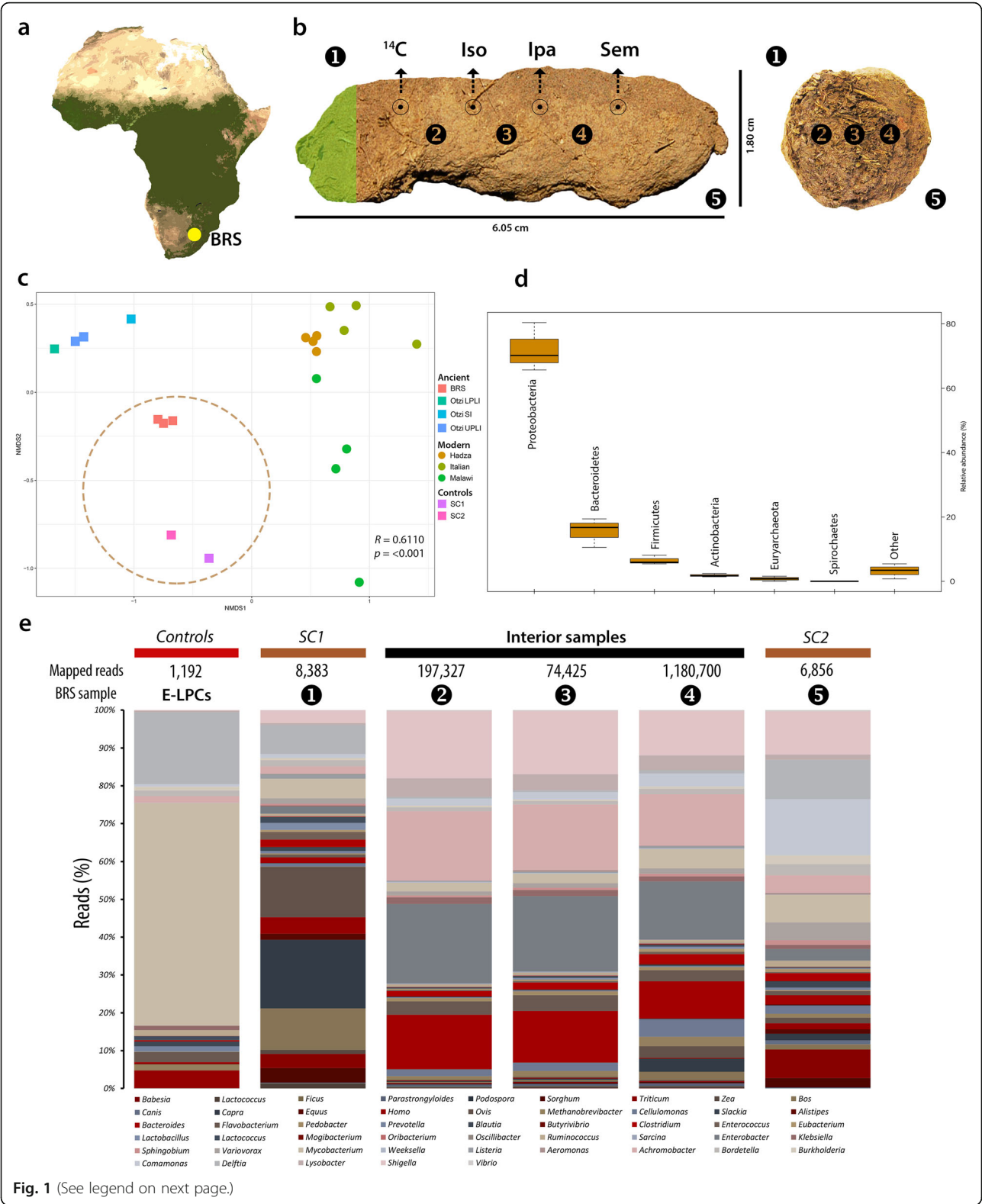
75 thousand years ago (kya) resulted in the global dispersal of our species [20]. Significantly, various microbes accompanied these human dispersals ‘out of Africa’ [21, 22]. Since the ancestral human IM is estimated to comprise a taxonomically and metabolically more diverse array of microbes than those found in contemporary societies [6, 10], the IMs of pre-Clovis North Americans [23], pre-Columbian Puerto Rican Amerindians [24] and pre-Columbian Andeans [25] represent more accurate indications of ancient human IM composition. These studies have provided significant insight into the structure, function and evolution of the human IM, highlighting the influence of dietary changes on the intestinal microbial ecology of contemporary humans [2, 6, 7]. These have also provided essential baseline information for understanding the evolutionary processes implicated in the taxonomic configuration and metabolic capacity of both healthy and dysbiotic human IMs.

Despite the fact that African populations are not underrepresented in studies concerning ‘traditional’ human IMs [8, 12, 26], there is, currently, no information concerning the taxonomic diversity and metabolic capacity of ‘ancestral’ (i.e. archaeologically-derived) African IMs. To gain insight into the ancient African human IM, the prehistoric incidence of intestinal parasites, pathogenic microbes and antibiotic resistance genes, we performed shotgun metagenomic sequencing of a prehistoric (pre-colonial) faecal specimen recovered from an Iron Age (c. AD 1460) context at Bushman Rock Shelter (BRS) in Limpopo Province, South Africa. Comparison with ancient (Ötzi), traditional (Hadza and Malawian) and contemporary ‘Western’ (Italian) IM datasets indicate that the IM of the BRS individual represents a unique taxonomic and metabolic configuration observed in neither contemporary African, nor European, populations (see the ‘Methods’ section).

## Results and discussion

### Specimen provenience and ancient subsistence reconstruction

The palaeo-faecal specimen was recovered in situ from an exposed stratigraphic section at BRS [27] (Fig. 1a, Fig. S1). This large dolomitic rock-shelter is situated on the edge of the Great Escarpment in the Drakensberg chain. The occupation level from which the specimen derives



(See figure on previous page.)

**Fig. 1** Provenience of sub-sampling protocol applied to and microbial taxa detected in the BRS palaeo-faecal specimen. **a** The location of Bushman Rock Shelter (BRS) in Limpopo Province, South Africa. **b** Lateral (left) and cross-sectional (right) views of the specimen indicating the sub-sampling protocol applied to facilitate DNA extraction, including ‘sedimentary control’ sample 1 (‘SC1’); faecal samples 2, 3 and 4 and sedimentary control sample 5 (SC2);  $^{14}\text{C}$  AMS dating ( $^{14}\text{C}$ ); isotope analyses (*Iso*); intestinal parasitic analyses (*Ipa*); scanning electron microscopy (*Sem*); and the preservation of a voucher sample (indicated in green shading). **c** Non-metric multi-dimensional scaling (NMDS) plot comparing the taxonomic community structure (by weighted Bray-Curtis dissimilarity analysis) of the BRS specimen (i.e. BRS2, BRS3 and BRS4) and the sediment controls (SC1 and SC2) with the ancient (Ötzi) (indicated as SI ‘small intestine’, LPLI ‘lower part of the lower intestine’ and UPLI ‘upper part of the lower intestine’), traditional (Hadza and Malawian) and modern (Italian) IM datasets (taxa were filtered for the occurrence of > 3 in at least 20% of the samples resulting in the inclusion of 371 taxa) ( $R = 0.6110$  indicates ANOSIM analysis which revealed significant differences ( $p = < 0.001$ ) between the ancient and modern IM samples). **d** Box-and-whisker plot indicating the relative abundance of intestinal bacterial phyla detected in the BRS specimen (i.e. BRS2, BRS3 and BRS4) (‘other’ comprises phyla with < 0.6% relative abundance). **e** Bar chart providing an overview of all environmental, commensal and pathogenic genera identified in the BRS specimen (BRS2, BRS3 and BRS4) and information concerning the DNA extraction and library preparation negative controls (E-LPCs) and modern and ancient sedimentary controls (SC1 and SC2) (data derived from Tables 1 and 2 and Table S1) (see the ‘Methods’ section)

comprises the uppermost archaeological unit of the rock-shelter, designated ‘Angel’ in ‘Block A’. This layer includes deeply incised ceramic fragments of the ‘Lydenburg’ cultural complex, as well as stone artefacts and material of recent (European) origin. Layer 1 (‘Angel’) dates to the arrival of Bantu-speaking Iron Age agro-pastoralists in the region after *c.* 1800 years ago (ya) (Fig. S1). This occupation therefore reflects the advent of the Neolithic in South Africa, which entailed the introduction of domesticated taxa such as sorghum (*Sorghum bicolor*), cattle (*Bos taurus*) and various other Iron Age-related species and cultural practices (e.g. ceramic and iron-smelting technologies) into the region [28]. All the preceding archaeological layers at BRS are representative of occupations by Holocene (e.g. the Oakhurst and Robberg techno-complexes ~ 10 kya and ~ 20 kya, respectively) and Pleistocene (i.e. the Pietersburg techno-complex ~ 70 kya) hunter-gatherers [27, 29].

Following recovery of the specimen using latex gloves and stainless steel forceps, whilst wearing a biologically impervious bodysuit and surgical face mask, it was sealed in a sterile plastic ziplock bag and stored at ~ 4 °C. Sediment control sample BRS1 (‘SC1’) was collected from the surface of the rock-shelter (~ 25 cm above the specimen) and BRS5 (‘SC2’) from the levels preceding the Iron Age (~ 25 cm below the specimen) in the Oakhurst occupation dated to *c.* 10 kya [27, 29]. These were used as ‘controls’ to assess ecological and faecal community composition for biological plausibility, and also the likelihood of sedimentary aDNA (sedaDNA) leaching. Sub-sampling was performed in ancient DNA (aDNA) laboratories at the Centre for GeoGenetics, University of Copenhagen (Denmark), using established protocols [30] (Fig. S2). Prior to sub-sampling, the outer surface or cortex (~ 5 mm) of the specimen was removed with a scalpel and excluded from further analyses, primarily as it was in contact with surrounding sediment, and could therefore have been contaminated by

soil-derived microbes (Fig. S2). To address within-sample variability, three faecal sub-samples (i.e. BRS2, BRS3 and BRS4) were taken from different sites within the specimen. From the remaining one-third of the specimen, four sub-samples were taken for radiocarbon ( $^{14}\text{C}$ ) dating, isotopic analysis and microscopic intestinal parasitic and scanning electron microscopy (SEM) analyses. One-sixth of the specimen was preserved (at – 20 °C) as a voucher sample (Fig. 1b).

To ascertain whether the palaeo-faecal specimen derives from a human individual, all other potential source species were systematically excluded, as per the following procedure. Given the limited number (1967) of aDNA sequence reads mapped to *H. sapiens*, likely due to the removal of the exterior cortex prior to sub-sampling (in which most human-derived (nuclear and mitochondrial) DNA would be expected to reside), metagenome assembly could not be performed. Note that, as discussed below, differences in taxonomic composition between the cores and cortices of ancient faecal specimens have been documented, with larger proportions of soil-derived taxa present in the cortices [24]. This provided the impetus for the removal of the faecal specimen’s cortex prior to sub-sampling and DNA extraction. Morphologically, the specimen resembles several candidate species [31], although no DNA sequence reads for indigenous felids (e.g. leopard (*Panthera pardus*), caracal (*Caracal caracal*) etc.), mustelids (honey badger (*Mellivora capensis*) and polecat (*Ictonyx striatus*)), jackal (*Canis mesomelas*) or domestic dogs (*Canis lupus familiaris*), and none for the indigenous primates, i.e. vervet monkeys (*Cercopithecus aethiops*) or baboon (*Papio ursinus*), were detected. Reads related to non-human primates, i.e. *Pan troglodytes* (common chimpanzee) and *Macaca mulatta* (rhesus macaque), are likely the result of false-positive identifications, as these taxa do not currently occur in the region, nor would they have in the past [32]. As discussed below, the presence



of specific authenticated ancient subsistence-related aDNA reads appears to be indicative of a diet typical of human individuals, and not of indigenous non-human omnivores. Similarly, the incidence of authenticated ancient intestinal taxa resembling the typical composition of the human IM, supports this conclusion. The incidence of statistically significant (i.e. verified ancient) C-T  $p$  values for the 1967 reads mapped to *H. sapiens* further supports the conclusion that the faecal specimen derives from a human individual (Table S1).

To confirm the association of the faecal specimen with the archaeological context from which it was recovered, two direct radiocarbon ( $^{14}\text{C}$ ) accelerator mass spectrometry (AMS) dates were generated from two sub-samples taken from within the specimen (Fig. 1b) (see the 'Methods' section). The calibrated dates of  $470 \pm 44$  years BP (IT-C-1020) and  $460 \pm 35$  years BP (IT-C-1077) indicate that the sample was deposited c. AD 1460, within the errors of radiocarbon dating (Table S2) (see the 'Methods' section) and with the calibration medians (cal. AD) of AD 1461 and AD 1464. This date falls within the South African Middle Iron Age (spanning AD 1300–1840), closely follows the demise of the nearby Kingdom of Zimbabwe at c. AD 1450 [33] and precedes first contact with European seafarers after AD 1488 [34].

Prior to the identification of environmental and subsistence-related taxa, all exotic taxa, including kiwi (*Apteryx*), carp (*Cyprinus*), salmon (*Oncorhynchus*), pig (*Sus*), chicken (*Gallus*) and rice (*Oryza*) were identified and excluded from further analyses. The evaluation of taxa present in the DNA extraction ( $n = 1$ ) and library preparation ( $n = 1$ ) negative controls (E-LPCs) indicated that instances of environmental contamination were restricted largely to taxa widely cited as either 'contaminants' or as derived from false-positive identifications [35, 36]. The authenticity of microbial and macrobial sequence-derived taxa was determined by statistical aDNA sequence damage estimation [37], comparison to E-LPCs, DNA read-length characteristics and ecological conformity. Using high-quality filtered reads for DNA damage estimation analyses with PMDtools [37], this process facilitated the validation of forty-seven taxa represented by 1,470,662 reads as 'ancient' (Fig. 2, Tables 1 and 2, Table S1) (see the 'Methods' section). Subject to the availability of sufficient numbers of high-quality 'mappable' aDNA sequence reads, we employed mapDamage [38, 39] to validate the authenticity of taxa by determining the incidence of C-T and G-A substitution rates at the 5'-ends and 3'-ends of strands (see the 'Methods' section). DNA damage analyses of sequence reads derived from the genera *Bacteroides* and *Shigella* (Fig. 2b, c) and also *Bos taurus* and *Sorghum bicolor* (Fig. 2d, e) indicate that the nucleotide composition at

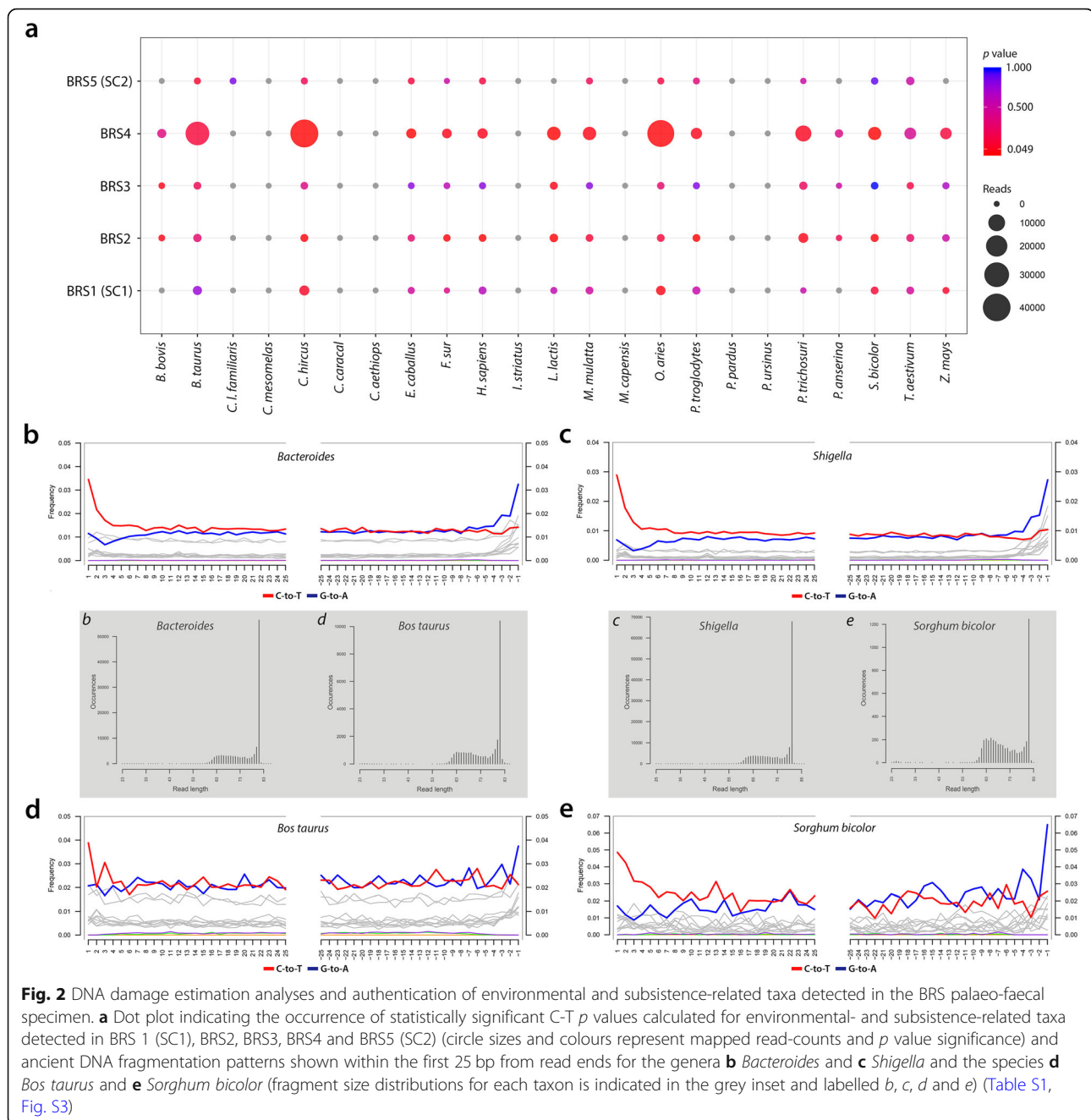
the ends of the analysed reads exhibits the typical pattern expected for ancient DNA [38, 39].

In dietary terms, the Bantu-speaker agro-pastoralist diet comprised not only domesticated animal and plant taxa, but also various hunted and gathered indigenous species, including antelope, fish, plants and fruits [40]. The presence of subsistence-related reads derived from sorghum (*S. bicolor*), cluster figs (*Ficus sur*), goat (*Capra hircus*), sheep (*Ovis aries*) and beef (*B. taurus*) is indicative of taxa that were consumed shortly before stool deposition by the BRS individual. Based on bulk untreated  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values obtained from isotopic analyses, the individual had a predominantly C4-based meal, with a minor C3-based contribution (see the 'Methods' section). This concurs with the aDNA evidence indicating the presence of sorghum, wild figs and beef. Sorghum is a C4 plant with published  $\delta^{13}\text{C}$  values of  $-14.0$  to  $-11.0\text{‰}$ , and cattle are grazers (i.e. C4 consumers). The  $\delta^{13}\text{C}$  values ( $-16.79\text{‰}$ ) are higher than  $-18\text{‰}$ , which is considered the threshold for a predominantly terrestrial diet. These values do not however preclude the occasional consumption of freshwater resources, including fish and freshwater mussels, given the close proximity ( $\sim 1.5$  km) of the shelter to the perennial Ohrigstad River [27] (Table S3, Fig. S4). The incidence of cattle (*B. taurus*), cattle-specific microbes, i.e. *Lactococcus lactis* (a component in fermented milks) and *Babesia bovis* (the causative agent of babesiosis) are also representative of a Bantu-speaker pastoralist subsistence economy. The non-authenticated (i.e. lacking the damage patterns typical of aDNA) incidence of *Podospora anserina* (a dung-colonising fungus) is interpreted as symptomatic of post-depositional saprophytic processes, likely resulting from the recent (i.e. historical) use of the rock shelter as a livestock enclosure.

#### Identifying commensal intestinal microbiota

It is estimated that the human IM harbours  $\sim 150$  to  $\sim 400$  species [41], most of which belong to the phyla Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria [6]. The variability of microbial taxonomic abundance, however, influences the identification of the common core IM [42]. As many microbes are capable of transient integration into the IM, where they influence the composition and metabolic activity of resident IM communities [43], essentially environmentally derived genera, i.e. *Bacillus*, *Dietzia*, *Microbacterium*, *Paracoccus*, *Pseudomonas*, *Staphylococcus* and *Streptomyces*, were omitted from further analyses. This was performed largely as a precautionary measure and does not preclude the occasional or variable occurrence of these taxa in the human IM [42, 43].

On the basis of taxa detectable at sequencing depth, metagenomic comparison of the shotgun reads with the



National Center for Biotechnology Information (NCBI) BLASTn non-redundant nucleotide (*nt*) database using MEGAN Community Edition (CE) v6.10.10, and the Burrows-Wheeler Aligner (BWA) facilitated the identification of 722,094 reads (1.48% of all sequence reads) representing thirty-six ancient commensal IM genera. Subsequent statistical DNA damage estimation resulted in the elimination of 29,012 sequence reads, derived from 12 non-authenticated bacterial genera, from the dataset, including *Bifidobacterium*, *Coprococcus*, *Dorea*, *Faecalibacterium*, *Mollicutes*, *Neisseria*, *Parabacteroides*,

*Phascolarctobacterium*, *Romboutsia*, *Roseburia*, *Ruminiclostridium*, *Tissierella* and *Treponema* (see the 'Methods' section) (Table S4). These taxa could not be authenticated as truly 'ancient', given the very low numbers of sequence reads generated. Since these have been found to inhabit contemporary human IMs, their respective low sequence read counts suggest that their limited detection in the ancient BRS faecal specimen likely reflects their corresponding depletion in the BRS IM. Based on the remaining authenticated (i.e. 'ancient') 688,089 sequence reads exhibiting an average read-length of

**Table 1** DNA sequence reads for twenty-four authenticated commensal IM taxa detected in the BRS palaeo-faecal specimen. Statistically significant (i.e. verified ancient) C-T *p* values are indicated in bold. BWA mapping was performed using high-quality filtered reads for DNA damage estimation analyses using PMDtools (C-T *p*-values) (see the 'Methods' section). Additional read-length information for individual taxa is provided in Table S5

Phylum	Genus	E-LPC reads	BWA aligned (mapped) sequence reads					C-T <i>p</i> values		Mean read length (bp)							
			BRS1 SC1	BRS2	BRS3	BRS4	BRS5 SC2	Total %	BRS IM total		BRS IM %						
Euryarchaeota	Methanobrevibacter	19	7	1952	1109	29147	69	32303	4.67	32208	4.68	70.60	–	0.193028900	0.073998830	<b>0.002632932</b>	0.567650800
	Actinobacteria	0	65	3445	1626	54666	143	59945	8.67	59737	8.68	66.60	0.921500700	<b>0.002823314</b>	0.153993200	<b>0.006118848</b>	0.509736300
Bacteroidetes	Slackia	0	5	89	29	1343	8	1474	0.21	1461	0.21	61.80	<b>0.007399490</b>	0.935197800	0.311623400	0.508960500	–
	Alistipes	2	4	139	46	1069	15	1275	0.18	1254	0.18	65.40	0.675706100	0.420822200	0.053055550	<b>0.031112450</b>	0.606658100
	Bacteroides	5	114	28316	10115	115446	162	154158	22.30	153877	22.36	72.20	0.721165000	<b>0.001696204</b>	<b>0.000000390</b>	<b>0.000622655</b>	0.142918300
	Flavobacterium	32	58	6986	3125	33452	72	43725	6.33	43563	6.33	64.40	0.216522900	<b>0.000235190</b>	<b>0.000039500</b>	<b>0.022810540</b>	0.921151200
	Pedobacter	1	23	1656	715	9663	19	12077	1.75	12034	1.75	62.80	0.500000000	<b>0.001704170</b>	0.075530930	0.062391830	0.161709000
	Prevotella	17	48	418	184	3712	43	4422	0.64	4314	0.63	61.00	0.099233850	0.305674400	0.081571890	<b>0.016625350</b>	0.631959800
Firmicutes	Blautia	15	72	276	124	3640	111	4238	0.61	4040	0.59	68.20	0.608740100	0.084446270	0.428864300	<b>0.045772080</b>	0.2456663300
	Butyrivibrio	0	8	325	127	3082	6	3548	0.51	3534	0.51	58.20	0.395839900	<b>0.037614750</b>	0.109365700	0.273060200	0.580047400
	Clostridium	4	148	2671	1304	29948	135	34210	4.95	33923	4.93	65.80	0.074740200	0.296300200	0.190844300	<b>0.015163840</b>	0.360285900
	Enterococcus	1	156	553	298	7427	20	8455	1.22	8278	1.20	74.60	<b>0.003765520</b>	0.333696100	<b>0.022915900</b>	0.274913800	–
	Eubacterium	0	40	697	296	9029	51	10113	1.46	10022	1.46	65.80	0.106194500	<b>0.004360991</b>	0.056443760	<b>0.024325710</b>	0.246347100
	Lactobacillus	0	143	0	259	6728	14	7144	1.03	6987	1.02	69.00	0.157217100	–	<b>0.011884140</b>	0.903384000	–
	Lactococcus	12	118	595	361	6219	14	7319	1.06	7175	1.04	69.60	0.463765800	<b>0.006836895</b>	<b>0.033206520</b>	<b>0.000000013</b>	–
	Mogibacterium	0	7	176	76	2116	5	2380	0.34	2368	0.34	68.60	0.499997800	0.228221200	<b>0.019403830</b>	<b>0.008329356</b>	0.188080100
	Oribacterium	1	12	170	72	2018	2	2275	0.33	2260	0.33	64.60	0.167621400	<b>0.050161290</b>	0.349725600	<b>0.000162173</b>	–
	Oscillibacter	0	7	165	91	1787	12	2062	0.30	2043	0.30	63.00	0.120518600	<b>0.009668940</b>	0.349725600	0.095004240	0.072742730
	Ruminococcus	18	42	1119	464	9229	100	10972	1.59	10812	1.57	67.40	0.271433200	<b>0.048583730</b>	0.738132500	<b>0.001960024</b>	0.300862800
	Sarcina	0	3	508	202	327	2	1042	0.15	1037	0.15	67.00	–	0.413740500	<b>0.018963150</b>	<b>0.025439260</b>	–
Proteobacteria	Enterobacter	0	145	41220	14798	181051	205	237419	34.34	237069	34.45	71.80	<b>0.004146302</b>	<b>0.000377506</b>	<b>0.000009640</b>	<b>0.001112472</b>	0.132896300
	Klebsiella	14	21	3485	1217	15633	72	20442	2.96	20335	2.96	70.00	0.529037900	<b>0.001153996</b>	<b>0.001264576</b>	<b>0.000436448</b>	0.921151200
	Sphingobium	0	36	1023	458	8330	84	9931	1.44	9811	1.43	65.80	0.126796900	0.132801100	<b>0.045433740</b>	<b>0.000753091</b>	0.602973600
	Variovorax	0	112	2099	878	16970	315	20374	2.95	19947	2.90	69.60	0.259295700	<b>0.005879342</b>	<b>0.057844450</b>	<b>0.020613870</b>	0.171048100
Total		141	1394	98083	37974	552032	1679	691303	100.00	688089	100.00	66.83	–	–	–	–	–



**Table 2** DNA sequence reads for twelve authenticated pathogenic taxa detected in the BRS palaeo-faecal specimen

Phylum	Genus	Pathogenic clinical manifestations	<i>E- LPC reads</i>	BWA aligned (mapped) sequence reads					Mean read length (bp)	C-T		<i>p</i> -value			
				BRS1 <i>SC1</i>	BRS2	BRS3	BRS4	BRS5 <i>SC2</i>		Total	BRS1 <i>SC1</i>		BRS2	BRS3	BRS4
Actinobacteria	<i>Mycobacterium</i>	Tuberculosis, leprosy, atypical infections	701	397	4535	1853	59096	487	67069	67.20	0.0707802	<b>0.0017862</b>	<b>0.0037356</b>	<b>0.0100530</b>	0.2953064
Bacteroidetes	<i>Weeksella</i>	Pneumonia, sepsis, peritonitis, urinary tract infections	1	0	461	175	2552	0	3189	45.40	–	<b>0.0069370</b>	<b>0.0484237</b>	0.1960360	–
Firmicutes	<i>Listeria</i>	Listeriosis (convulsions, septicaemia and meningitis)	0	92	453	233	5989	5	6772	72.80	<b>0.0000005</b>	0.3494000	0.0975000	0.8800000	–
Proteobacteria	<i>Aeromonas</i>	Gastroenteritis and wound infections	0	13	129	289	2033	29	2493	67.20	0.2264910	<b>0.0206110</b>	0.4139431	<b>0.0060921</b>	0.4998525
	<i>Achromobacter</i>	Skin, soft tissue and respiratory infections	21	150	36089	12869	160087	311	209527	72.40	<b>0.0011408</b>	<b>0.0017379</b>	<b>0.0000094</b>	<b>0.0015504</b>	0.1190260
	<i>Bordetella</i>	Pertussis (whooping cough, respiratory infection)	18	123	2027	683	16466	196	19513	69.60	<b>0.0204611</b>	0.1325985	<b>0.0082522</b>	<b>0.0077918</b>	0.2557714
	<i>Burkholderia</i>	Pulmonary infections (pneumonia, melioidosis)	12	48	869	287	6965	156	8337	62.00	0.2711432	<b>0.0462580</b>	0.2140947	0.3505477	0.1594083
	<i>Comamonas</i>	Acute appendicitis, urinary tract infection	8	79	3656	1467	40820	987	47017	73.40	0.6379285	<b>0.0059303</b>	<b>0.0225115</b>	<b>0.0322217</b>	0.6381353
	<i>Delftia</i>	Endocarditis, skin an ocular infections	226	594	921	348	10552	692	13333	74.80	0.2733595	<b>0.0035008</b>	0.5813131	0.0712462	0.0567663
	<i>Lysobacter</i>	Opportunistic pathogen, various symptoms	5	37	9616	3170	45813	95	58736	66.20	0.2478419	<b>0.0013587</b>	<b>0.0003318</b>	<b>0.0064930</b>	0.8210803
	<i>Shigella</i>	Shigellosis (dysentery, seizures, mucosal ulceration)	2	253	35046	12402	137623	762	186088	76.80	0.5381291	<b>0.0002471</b>	<b>0.0000002</b>	<b>0.0002182</b>	0.2954809
	<i>Vibrio</i>	Cholera (gastroenteritis, diarrhoea, septicaemia)	0	7	419	147	2333	21	2927	62.80	–	0.4420414	0.1131828	<b>0.0005438</b>	–
<i>Total</i>			<b>994</b>	<b>1793</b>	<b>94221</b>	<b>33923</b>	<b>490329</b>	<b>3741</b>	<b>625001</b>	<b>67.55</b>	–	–	–	–	–

Statistically significant (i.e. verified ancient) C-T *p* values are indicated in bold text. BWA mapping was performed using high-quality filtered reads for DNA damage estimation analyses using PMDtools (C-T *p* values') (see the 'Methods' section). Additional read-length information for individual taxa is provided in [Table S5](#)

66.83 base pairs (bp), twenty-four ancient commensal IM taxa (Table 1) were identified. It is of interest to note that, whereas the BRS1 ‘surface control’ sample (SC1) yielded authenticated reads derived from ancient microbial IM taxa (i.e. *Enterobacter*, *Enterococcus* and *Slackia*), the much older BRS5 control (SC2) (i.e. the Oakhurst occupation dated to c. 10 kya) did not (Tables 1 and 2, Table S1). The IM of the BRS individual, including only the interior sub-samples that is BRS2, BRS3 and BRS4, was determined to be dominated by the phyla Proteobacteria (41.73%) and Bacteroidetes (31.25 %), followed by Firmicutes (13.44%), Actinobacteria (8.89%) and Euryarchaeota (4.68%) (Table 1, Fig. 1d). At genus level, the bulk (71.82%) of reads was ascribed to *Enterobacter* (34.45%), *Bacteroides* (22.36%), *Cellulomonas* (8.68%) and *Flavobacterium* (6.33%). In addition to *Clostridium* (4.93%) and *Methanobrevibacter* (4.68%), all other genera exhibit < 5% relative abundance. We note that the use of ‘relative abundance’ as a measure of taxonomic representation has been a standard means by which differences in taxonomic composition or ‘abundance’ in IM datasets is analysed, verified and compared, and that various notable IM studies have adhered to the use of ‘relative abundance’ as standard analytical protocol [6, 8, 9, 12, 24, 25, 44]. In addition, while cognisant of the compositional complexity of microbiome samples [44] and the possible influence of the fragmented nature of ancient microbial DNA on taxonomic classification [45], we note that ancient DNA damage has been revealed to exert a minimal influence on species detection and on the ‘relative abundance’ of IM taxa in both simulated ancient and modern datasets [46].

#### Identification of ancient pathogenic microbial taxa

There is an estimated 1407 recognised species of human pathogens [47], many of which influence not only health and immune responses [48], but also cognitive development [49] and social behaviour [50]. On the basis of taxa detectable at sequencing depth, metagenomic comparison of the shotgun reads with the NCBI BLASTn non-redundant nucleotide (nt) database using MEGAN and BWA, facilitated the identification of 625,001 ‘mappable’ reads (1.34% of all sequence reads) representing twelve ancient potentially pathogenic taxa (Table 2). Only authenticated ancient taxa were retained (exhibiting an average read-length of 67.55 bp), the authenticity of which was determined by assessing the incidence of statistically significant ( $p = < 0.05$ ) C-T substitutions at the 5′-ends of sequence reads (Table 2).

The occurrence of authenticated ancient reads homologous to *Listeria* and restricted to BRS1 (SC1) suggests that this taxon, although ancient, is most likely environmental (i.e. sedimentary) and that it does not derive from the faecal specimen. Conversely, the incidence of

authenticated DNA reads for *Mycobacterium* in BRS2, BRS3 and BRS4, and not in SC1 or SC2, is indicative of the ancient intestinal presence of this genus, although it is also commonly found in environmental (i.e. sedimentary) contexts. While *Mycobacterium* is not a known member of the intestinal microbiota, and given that not all species are pathogenic, the presence of authenticated ‘ancient’ reads derived from *Mycobacterium* within the faecal specimen can neither be confirmed, nor precluded, as symptomatic of an infection. Some species, particularly *M. avium*, is known to invade intestinal epithelial cells and has been implicated in ulcerative colitis [51]. Similarly, the incidence of authenticated reads for the genera *Comamonas*, *Lysobacter* and *Shigella* in BRS2, BRS3 and BRS4, *Aeromonas* in BRS2 and BRS4, and *Vibrio* in BRS4 confirms that these taxa derive from the faecal specimen. Given that not all species of these genera are pathogenic, their presence is not necessarily symptomatic of an ancient gastro-intestinal infection. *Achromobacter* occurs in all sub-samples except for the ancient (c. 10 kya) control (BRS5/SC2) which did not yield any authenticated microbial pathogenic taxa. The notable abundance of commensal (41.73 %) (Table 1) and pathogenic (87.67%) (Table 2) members of the Proteobacteria in the BRS IM are of interest as it has been established that an increase in Proteobacteria is indicative of IM dysbiosis and metabolic disease [52]. Compared to primary human IM phyla, i.e. Actinobacteria, Bacteroidetes and Firmicutes, the relative abundance of Proteobacteria in the IM is, however, highly variable. While an increase in the abundance of Proteobacteria, especially members of the *Enterobacteriaceae* (i.e. *Klebsiella*, *Salmonella* and *Shigella*) [53] is a feature of bacterial dysbiosis, the human IM also contains members of commensal Proteobacteria, i.e. *Enterobacter*, *Klebsiella*, *Sphingobium* and *Variovorax*. Under ‘healthy’ conditions, the relative abundance of Proteobacteria in the human IM can increase to ~ 45% without observable clinical implications [51].

Microscopic analysis aimed to determine the presence of intestinal parasites, namely helminths and protozoa, did not yield conclusive results as no physical evidence (e.g. eggs and larvae) of an intestinal parasitic infection could be found (see the ‘Methods’ section). Although this concurs with the aDNA results, the analyses of a single sub-sample might not have been sufficient to detect intestinal parasitic remains. Conversely, not all members of a population would necessarily be infected by intestinal parasites, possibly because of either natural resistance or limited exposure to contaminant sources. Similarly, while SEM analyses did not result in the detection of parasitic remains, it did facilitate the detection of desiccated bacterial cells, degraded plant fragments and plant stellate hair (Fig. S5).

### Ancient and modern IM taxonomic comparisons

In terms of taxonomic composition, the ancient samples (BRS and Ötzi) exhibit spatial separation from the ‘traditional’ (Hadza, Malawian) and modern (Italian) comparative cohorts. Our selection of the ancient Alpine (Tyrolean) Iceman (Ötzi) as a comparative IM dataset relates to his prominence as a well-studied individual representative of an ancient, i.e. authenticated and securely dated pre-industrial, human IM. As there is currently no comparatively ancient datasets available from indigenous African contexts, we selected available IM datasets of contemporary ‘traditional’ African (i.e. Hadza and Malawian) and also one contemporary European (i.e. Italian) comparative dataset.

Hierarchical clustering using complete-linkage based on Spearman’s correlations, produced a clear separation of ancient and modern populations. ANOSIM analysis revealed significant differences between the ancient and modern IM samples ( $R = 0.9098$ ,  $p = < 0.001$ ) for 371 taxa (Fig. S6) (see the ‘Methods’ section). As stated, and bearing in mind the compositional complexity of IM samples [44] and the conceivable influence of fragmented DNA on taxonomic classification [45], ancient DNA damage results in minor differences in species detection and in comparisons concerning the ‘relative abundance’ of microbial taxa identified in ancient and modern IM datasets [46]. As with weighted Bray-Curtis analysis based on the relative abundance of all identified IM taxa (Fig. 1c), we note that unweighted Bray-Curtis analyses based on the ‘presence-absence’ of IM taxa exhibits correspondingly evident differences between the sedimentary controls (SC1 and SC2) and the ancient (i.e. BRS and Ötzi), ‘traditional’ (Hadza, Malawian) and modern (Italian) comparative IM cohorts (ANOSIM  $R = 0.8361$ ,  $p = < 0.001$ ) (Fig. S7). With regards potential contamination derived from the surrounding archaeological sedimentary matrix in the BRS palaeo-faecal specimen, comparison of the incidence of the twenty-four authenticated ancient IM taxa (Table 1) indicates that the surrounding sedimentary matrix (BRS1, ‘SC1’ and BRS5, ‘SC2’) and the DNA extraction and library preparation negative controls (E-LPCs) are not significant sources of microbial taxa identified in the palaeo-faecal specimen (BRS2, BRS3 and BRS4) (Fig. S8).

Metagenomic comparison of all analysed shotgun reads revealed that the IM of the BRS individual (i.e. BRS2, BRS3 and BRS4) is characterised by a Firmicutes/Bacteroidetes (F/B) ratio significantly skewed towards Bacteroidetes (at 31.25%), as opposed to Firmicutes (at 13.44%) (Fig. 1d, Table 1). The F/B ratio [54] is widely considered as significant in human IM composition, with dysbiosis associated with inflammation, obesity and metabolic diseases [55]. Although this significance is controversial [56], we note that the BRS F/B ratio (0.4)

does not resemble those reported for modern ‘traditional’ Bantu-speaking Africans in Burkina Faso (2.8) [57] nor that calculated here for the East African Hadza (2.6) [8]. This can likely be attributed to the fact that ‘traditional’ diets rich in starches (e.g. potatoes, yams and sweet potatoes) have been shown to increase the F/B ratio, including increases in relative abundance of Firmicutes and enzymatic pathways and metabolites involved in lipid metabolism [58].

The IMs of modern humans have furthermore been stated to comprise one of three ‘enterotypes’, based on prevailing genera, i.e. *Bacteroides*, *Prevotella* or *Ruminococcus* [59]. Some taxa relate to long-term diets, such as *Bacteroides*, which is associated with the consumption of large amounts of protein and animal fat, and *Prevotella*, which is indicative of a high plant-derived carbohydrate intake [60]. Similarly, *Ruminococcus* prevails in individuals who consume significant amounts of polyunsaturated fats, e.g. marine fish, vegetable oils and nuts and seeds. The enterotypic composition of the BRS IM diverges from that reported for ‘traditional’ Africans [8, 12, 15, 26, 57, 61]. In relation to *Ruminococcus* (1.57%) and *Prevotella* (0.63%), the BRS IM is characterised by a predominance of *Bacteroides* (22.36%) (i.e. ‘Enterotype 1’) which concurs with a diet rich in protein and animal fat, and which lends support our interpretation of the BRS *Bacteroidetes*-dominated F/B ratio. While this corresponds to data reported for the West African BaAka [12], it differs from the IM taxonomic composition reported for modern African cohorts, including the Tanzanian Hadza [8] and children in Burkina Faso [57], which exhibits higher abundance of *Prevotella* (Fig. 1e). The sizable incidence of *Flavobacterium* (6.33%) in the BRS IM likely relates to the fact that members of this genus are resistant to dietary phenolic compounds derived from largely ‘medicinal’ plant taxa, including phenolic acids, flavonoids, tannins, curcuminoids, coumarins, lignans and quinones [62]. This genus also occurs in the IMs of non-human primates, including baboons (*P. ursinus*) and gorillas (*Gorilla gorilla*) [63]. In relation to its substantial presence in the BRS IM, members of this genus might also have played a role in the elimination of aflatoxins present in milk, cheese, grains and figs. *Methanobrevibacter* (4.68%) is the most abundant archaeon in the human IM [64]. Besides consuming fermentation products produced by saccharolytic bacteria, archaeal methanogenesis also improves the efficiency of polysaccharide fermentation.

The BRS IM furthermore exhibits enrichment towards *Cellulomonas* (8.68%) which degrades cellulose [65], *Clostridium* (4.93%) which is essential for IM resistance to infection and dysbiosis [66] and *Pedobacter* (1.75%) and *Prevotella* (0.63%) which, resembling non-pathogenic *Treponema*, are cellulose and xylan

hydrolyzers [8]. *Alistipes* (0.18%) is associated with a protein-rich diet and involved in amino acid fermentation [61], and *Butyrivibrio* (0.51%) ferments sugars, cellodextrins and cellulose [67]. Since the antiquity of *Treponema* (*Spirochaetes*) in the BRS IM could not be verified, we cannot substantiate the premise that *Treponema* is inherently characteristic of all ‘traditional’ IMs [8, 9]. Less abundant taxa, i.e. *Ruminococcus* (1.57%), *Eubacterium* (1.46%) and *Enterococcus* (1.20%), are implicated in the digestion of starches and vitamin synthesis. *Sarcina* (0.15%) synthesises microbial cellulose and occurs in high numbers amongst yam-farming African Pygmy hunter-gatherers, and traditional populations in Papua New Guinea [68].

#### Ancient and modern IM metabolic comparisons

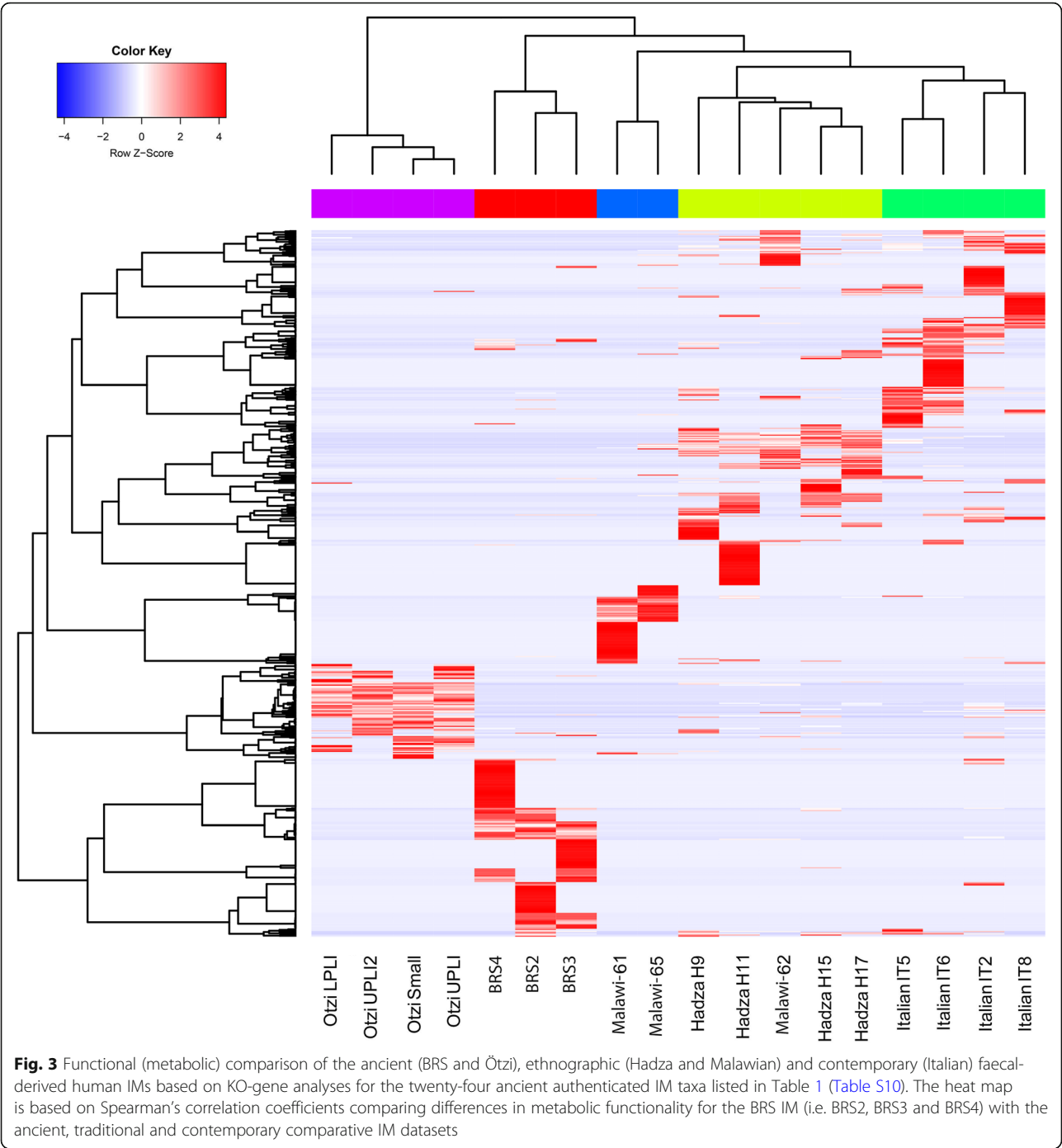
Despite having highly divergent IM taxonomic compositions, functional gene profiles are relatively similar amongst different contemporary human populations. Accordingly, microbial community (i.e. taxonomic) composition does not afford a thorough understanding of microbial IM community function (i.e. metabolic capacity) [69]. To ascertain statistically significant differences between the metabolic capacities of ancient (pre-industrial) and modern ‘Westernised’ human IMs, we explored and compared the functional IM capacities of two ancient (BRS and Ötzi), two traditional (Hadza and Malawian) and one modern ‘Western’ (Italian) human cohorts (see the ‘Methods’ section). This was performed only for the twenty-four ancient authenticated IM taxa indicated in Table 1. ANOSIM analysis revealed significant difference ( $R = 0.5395$ ,  $p = 0.001$ ) in the metabolic capacity of twenty-four authenticated ancient taxa for the ancient and modern cohorts. Spearman’s correlation was performed on the taxa linked to the KEGG categories (i.e. 1487 KO gene categories linked to specific IM taxa). Although not contiguous to the faecal specimen, the considerable differences in the incidence of particular KEGG Orthology (KO) genes in BRS1 (SC1) and BRS5 (SC2) (the younger surface-derived and the ancient Oakhurst (c. 10 kya) sediment samples), and BRS2, BRS3 and BRS4 eliminate the sedimentary matrix as a source for the greater proportion of KO genes identified in the faecal specimen (Fig. 3). The differential distribution of the 24 authenticated ancient IM taxonomic categories (Table 1), particularly in terms of the taxa detected in SC1 and SC2 vs. those detected in the BRS specimen (BRS2, BRS3 and BRS4) ( $R = 0.8361$ ,  $p = < 0.001$ ), lends support to this conclusion (Fig. S7).

ANOSIM analysis revealed significant differences ( $R = 0.4840$ ,  $p = < 0.001$ ) in the metabolic capacity of the ancient and modern comparative cohorts (Fig. 4a). Based on the analyses of the KO gene categories occurring only in the twenty-four authenticated ancient taxa (i.e. 1487)

(Table 1), 72 taxa-specific KO genes are identified as unique to the BRS IM (Fig. 4b). Metagenomic comparison of the shotgun reads with the BLASTx NCBI non redundant protein (*nr*) database using DIAMOND v0.8.36.98, and MEGAN CE v6.10.10 revealed that 117 taxa-specific KO genes (7.86%) are shared between all (i.e. ancient and modern) IM cohorts. While this is indicative of the relative temporal stability of a core commensal human IM community, it also reflects the variable (i.e. adaptable) and transient nature of human commensal IM community composition and metabolic capacity [3, 39, 70]. This responsive adaptability is echoed by the variable co-abundance of several relevant metabolic pathways identified in the BRS IM and the ancient and modern cohorts (Fig. 4c, Tables S6, S7, S8, S9).

In our comparison of the functional IM capacities of the two ancient, two traditional and one modern ‘Western’ human IM cohorts, we also focussed our analyses only on the twenty-four authenticated ancient IM taxa indicated in Table 1 (see also Table S8, S9, S10). In relation to the modern (Italian, Hadza and Malawian) and ancient (Ötzi) datasets, the BRS IM (i.e. BRS2, BRS3 and BRS4) exhibits enrichment of KO genes implicated in the biosynthesis of secondary metabolites, including K00163 (Kruskal-Wallis value ( $H$ ) = 14.151 and  $p$  value ( $p$ ) = 0.002), K00164 ( $H$  = 14.096,  $p$  = 0.002), K00163 ( $H$  = 15.812,  $p$  = 0.014), K00600 ( $H$  = 11.243,  $p$  = 0.004), K00568 ( $H$  = 11.706,  $p$  = 0.003) and K00457 ( $H$  = 14.762,  $p$  = 0.002) (Table S10). K00568 and K00457 are also implicated in the biosynthesis of terpenoid quinones. The capacity to biosynthesise toxic secondary metabolites (e.g. polyketides, isoprenoids, aromatics (phenylpropanoids) or alkaloids) is essential when dietary sources comprise largely natural and unprocessed foods. The BRS IM also exhibits enrichment of genes implicated in glyoxylate and dicarboxylate metabolism (K03781:  $H$  = 15.275,  $p$  = 0.018 and K00600) and the citric acid cycle (CAC) (K00164:  $H$  = 15.650,  $q$  = 0.015 and K02274:  $H$  = 15.928,  $p$  = 0.014). Glyoxylate and dicarboxylate glyoxylate are involved in the biosynthesis of carbohydrates, and CAC facilitates the release of energy from dietary carbohydrates, proteins and fats. Genes involved in glycolysis and gluconeogenesis (K00163:  $H$  = 15.812,  $p$  = 0.015) (pyruvate dehydrogenase E1 component) are also enriched.

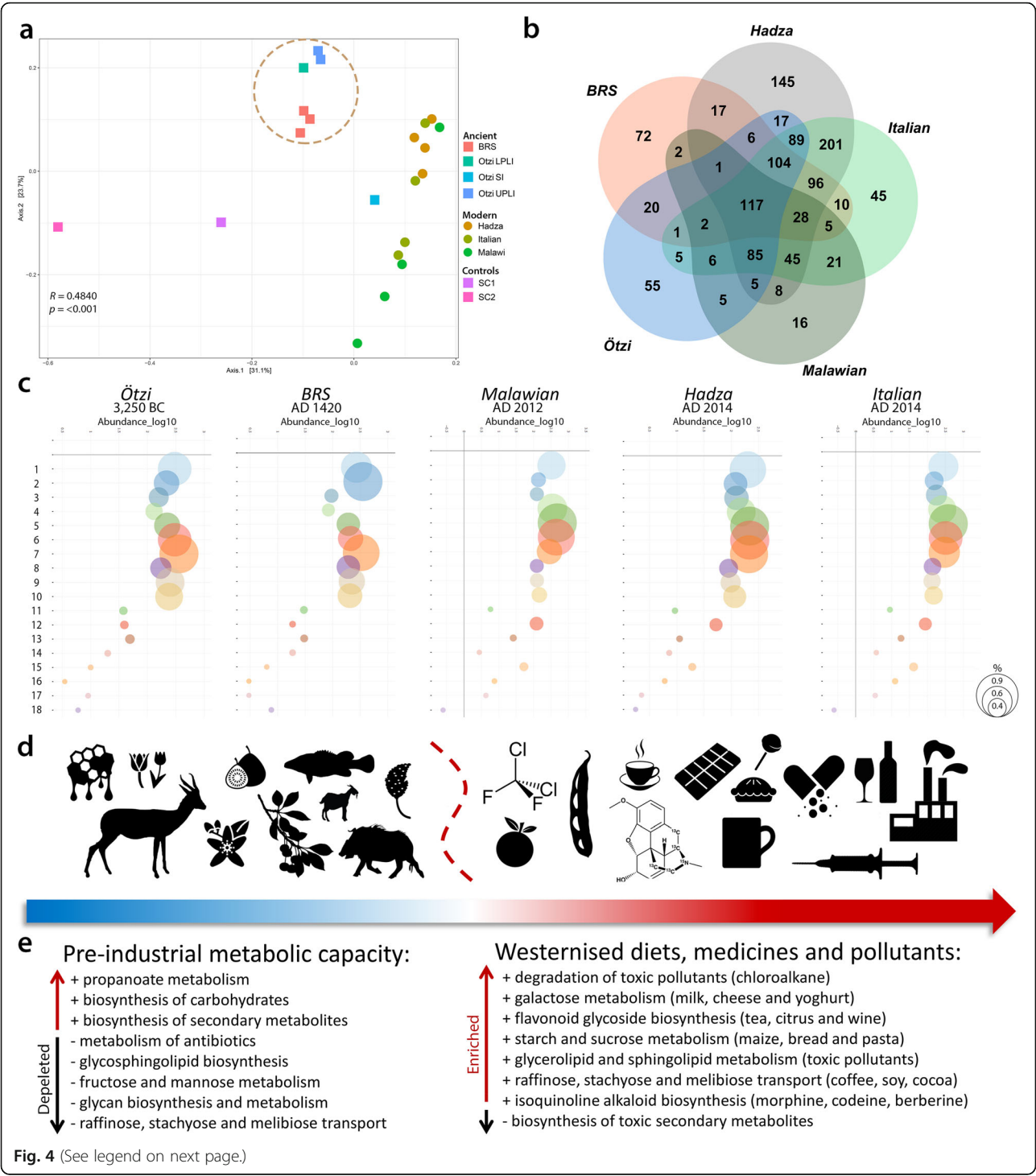
The BRS IM exhibits depletion of KO genes involved in raffinose, stachyose and melibiose transport (e.g. K10119:  $H$  = 15.934,  $p$  = 0.015; K10118:  $H$  = 15.640,  $p$  = 0.015; and K10117:  $H$  = 16.383,  $p$  = 0.011) (Table S8). Soybeans are primary dietary sources of raffinose and stachyose, and melibiose occurs in coffee, cacao and processed soy [71]. Deglycosylation by intestinal epithelial cell beta-glucosidases is a critical step in the



metabolism of dietary flavonoid glycosides derived specifically from tea, citrus and wine (K05349:  $H = 15.068$ ,  $p = 0.019$ ). The BRS IM also contrasts with the modern cohort in terms of the depletion of genes involved in glycan (sugar-chain) biosynthesis and metabolism (alpha-mannosidase) (K01206:  $H = 14.443$ ,  $p = 0.025$ ) (beta-galactosidase) (K01190:  $H = 15.777$ ,  $p = 0.015$ ). While modern oligosaccharides are derived largely from processed ‘table sugar’ comprising mainly sucrose and

fructose, foremost natural sources of sugar, comprising fruits and honey, would not have been consistently available for consumption. Correspondingly, KO genes involved in starch and sucrose metabolism (K00705:  $H = 15.318$ ,  $q = 0.018$  and K00975:  $H = 15.438$ ,  $p = 0.017$ ) (starch phosphorylase) (K00688:  $H = 13.496$ ,  $p = 0.035$ ) and fructose and mannose metabolism (K01193:  $H = 14.468$ ,  $p = 0.025$ ) (6-phosphofructokinase 1) (K00847:  $H = 16.009$ ,  $p = 0.014$ ) (fructokinase) are also depleted.





(See figure on previous page.)

**Fig. 4** Graphic summary of dietary and environmentally induced differences in the metabolic capacities of the ancient and modern IM datasets analysed in this study. **a** Principal coordinates analysis (PCoA) and comparison of the metabolic (functional) capacity of the BRS specimen (i.e. BRS2, BRS3 and BRS4) and the sediment controls (SC1 and SC2) with the ancient (Ötzi) (SI 'small intestine', LPLI 'lower part of the lower intestine' and UPLI 'upper part of the lower intestine'), traditional (Hadza and Malawian) and modern (Italian) IM datasets (KEGG categories were filtered for occurrence of > 3 in at least 20% of the samples). **b** Venn diagram indicating the relative abundance of IM taxa-linked KO genes identified in the ancient, traditional and modern comparative cohorts, calculated as based on the twenty-four authenticated ancient IM taxa indicated in Table 1. **c** Bubble charts indicating the co-abundance ( $\log_{10}$ ) of eighteen (labelled '1' to '18') metabolic IM capacities for the ancient, traditional and modern IM cohorts (bubble sizes are representative of the relative abundance of KEGG categories (see scale on right) and comprise (1) glycolysis/gluconeogenesis, (2) citrate cycle, (3) fructose/mannose metabolism, (4) galactose metabolism, (5) starch/sucrose metabolism, (6) amino sugar and nucleotide sugar metabolism, (7) pyruvate metabolism, (8) glyoxylate/dicarboxylate metabolism, (9) propanoate metabolism, (10) butanoate metabolism, (11) synthesis and degradation of ketone bodies, (12) sphingolipid metabolism, (13) biosynthesis of unsaturated fatty acids, (14) n-glycan biosynthesis, (15) glycosphingolipid biosynthesis (-globo), (16) glycosphingolipid biosynthesis (-ganglio), (17) chloroalkane/chloroalkene degradation and (18) naphthalene degradation) (Table S7). **d** Dissimilarities in ancient and modern IM metabolic capacities are related to recent (historical) changes in human dietary composition and exposure to toxic environmental pollutants (as indicated by the icons and the blue and red arrow). **e** Differences in IM metabolic capacities are contrasted in terms of the up- and down-regulation of IM metabolic capacities as an 'ancient' vs. 'modern' comparative summary (see the 'Methods' section)

While it appears that the depletion of KO genes related to fructose and mannose metabolism can be attributed to the depletion of bacterial taxa harbouring these genes, we cannot exclude the possibility that this might also be due to the differential preservation of aDNA derived from these bacterial taxa.

Whereas the Hadza IM is enriched in genes involved in fructose and mannose metabolism, the Italian IM is enriched in genes involved in the metabolism of simple sugars, e.g. glucose, galactose and sucrose [26]. KO genes involved in the degradation of toxic pollutants (i.e. chloroalkane, chloroalkene and naphthalene) and butanoate metabolism (K04072:  $H = 13.468$ ,  $p = 0.036$  and K00128:  $H = 15.970$ ,  $p = 0.013$ ) are also depleted in the BRS IM. This is noteworthy, as there are no known natural sources of chlorinated paraffins (CPs) [72]. CPs, including chloro-alkanes ( $C_{10-13}$ ), are widely used in the production of refrigerants, solvents, plasticisers and fire-retarding agents [73]. Naphthalene ( $C_{10}H_8$ ), a polycyclic aromatic hydrocarbon, is derived from petroleum distillation and is used in the manufacture of plastics, resins, fuels and insecticides. In addition, the depletion of KO genes involved in galactose, glycerolipid and sphingolipid metabolism and glycosphingolipid biosynthesis (K07407:  $H = 14.352$ ,  $p = 0.026$ ) (alpha-galactosidase) (K01190) (beta-galactosidase) is significant as it provides insight into the influence of exposure to modern environmental pollutants on IM composition and metabolic capacity.

Although it is challenging to infer ancient diet from ancient IM data, there is a growing understanding of the role of dietary choices on IM composition [2]. The enrichment, in the BRS IM, of genes serving specific metabolic processes (i.e. the biosynthesis of secondary metabolites, xenobiotic biodegradation, the metabolism of terpenoids and polyketides, propionate and butanoate and lysine degradation and the synthesis of ketone bodies) is noteworthy as it suggests that the BRS

metabolic profile is indicative of a diet rich in unprocessed natural resources, conceivably comprising medicinal plant substances (given the capacity for biosynthesising secondary metabolites and biodegrading xenobiotics), and encompassing irregular dietary intake. In addition, and given the enrichment of KO genes implicated in the synthesis and degradation of ketone bodies (i.e. K00626), the metabolic profile of the BRS individual approximates that induced by a ketogenic diet [74], characterised by high-fat, adequate-protein and low-carbohydrate dietary consumption, and accompanied by prolonged exercise and periods of low dietary intake or unintentional 'fasting' (Fig. 4d, Table S9). Whether this metabolic profile resembles that generally referred to as a 'palaeo-diet' is unclear, as this would have entailed the exclusion of dairy, grains and legumes, nutritional categories which did indeed form part of the BRS diet. The BRS IM, and also that of Ötzi, are furthermore characterised by depletion of KO genes involved in the metabolism of antibiotics (e.g. K11358:  $H = 15.320$ ,  $p = 0.017$ ), including aminocoumarin antibiotics, and the metabolism of isoquinoline alkaloids, including the opiates morphine and codeine, as well as the antibiotic berberine.

In contrast to the ancient IMs, the modern (Hadza, Malawian and Italian) IMs are characterised by enrichment of KO genes involved in raffinose, stachyose and melibiose transport (e.g. K10119, K10118 and K10117) indicative of a diet comprising soy, coffee, cacao and dietary flavonoid glycosides derived from tea, citrus and wine (K05349) (Table S10). This group also exhibits enrichment in genes concerning galactose metabolism, i.e. K07407 ( $H = 14.352$ ,  $p = 0.0259$ ) (alpha-galactosidase), K00849 ( $H = 15.553$ ,  $p = 0.016$ ) (galactokinase) and K00965 ( $H = 15.694$ ,  $p = 0.015$ ) (UDPglucose-hexose-1-phosphate uridylyltransferase). Galactose is metabolised from lactose (milk sugar consisting of disaccharide

glucose and galactose), the primary dietary source of which is milk and yoghurt. KO genes involved in starch and sucrose (K00705), and amino sugar and nucleotide sugar metabolism (K00965 and K00849:  $H = 15.553$ ,  $p = 0.016$ ) are also enriched, as are genes involved in glycine, serine, threonine, methane and antibiotic metabolism. The enrichment of genes involved in glycerolipid and sphingolipid metabolism, and glycosphingolipid biosynthesis (e.g. K01190 and K07407) (alpha-galactosidase), likely reflects the impact of modern environmental pollutants on IM composition and metabolic capacity. It would be of interest to determine whether this does in fact represent a 'population-wide' functional response to exposure to toxic compounds ubiquitous in modern industrialised urban environments.

In summary, significant differences between the ancient and modern IM metabolic capacity comprise the ability of the modern IMs to metabolise opiates (i.e. morphine and codeine), antibiotics (i.e. berberine), raffinose, stachyose and melibiose indicative of a diet comprising soy, coffee, cacao and dietary flavonoid glycosides derived from tea, citrus and wine and glycerolipid and sphingolipid metabolism. Since these compounds were not present at the time of deposition of the BRS specimen, our results document the pervasive evolutionary influence of dietary changes, medicinal treatments and environmental pollutants on the IM taxonomic composition and metabolic capacity of contemporary human populations (Fig. 4e, Table S8, S9).

### Antibiotic resistance genes

Our results also confirm reports of antibiotic resistance genes (ARGs) previously recovered from ethnographic cohorts and archaeological faecal samples [5, 25, 75]. Following analysis of the ancient and modern resistomes using Resistome Analyser (<https://github.com/cdean/resistomeanalyser>) (see the 'Methods' section), we identified a total of 15 functional ARGs, four of which occur in the BRS IM. These include the prokaryotic protein synthesis elongation factor Tu (EF-Tu) (*tufA* and *tufB*), fluoro-quinolone-resistant DNA topoisomerase (*parE*) and daptomycin-resistant *rpoB* (Table S11, Fig. S9). Note that ARGs were identified only in BRS4, and not in BRS1, BRS2 or BRS3, and is likely attributable to the variable sequencing depth of the Illumina platform used (see Supplementary Table 12). Several bacterial taxa (e.g. *Escherichia coli*, *Staphylococcus aureus* and *Streptomyces collinus*) have a duplicate of genes encoding EF-Tu (*tufA* and *tufB*) which confers resistance to the antibiotic kirromycin. These genes are also present in the Ötzi (sample UPLI), Hadza (H9 and H11) and Italian (IT6) datasets. In *Pseudomonas aeruginosa* and certain *Salmonella* serovars, *parE* confers resistance to fluoroquinolones, and in *Staphylococcus aureus* it confers

resistance to fluoroquinolones and aminocoumarin. These genes are present in Ötzi (UPLI) and the Hadza (H11). The daptomycin-resistant *rpoB* gene encodes the  $\beta$  subunit bacterial RNA polymerase and is the site of mutations that confer resistance to the daptomycin antibacterial agents, which is synthesised by *S. roseosporus*. Daptomycin-resistant *rpoB* is present in the Ötzi (SI, LPLI and UPLI) and Hadza (H11) datasets, but absent from the Malawian and Italian cohorts. Certain mutations in the RNA polymerase  $\beta$  subunit have been found to reduce the susceptibility of methicillin-resistant *S. aureus* (MRSA) for the antibiotics daptomycin and vancomycin [76]. A number of ARGs are limited in occurrence to the modern comparative cohorts. Several ARGs, including *gyrA*, which confers fluoroquinolone resistance to *Neisseria gonorrhoeae* and *Ureaplasma urealyticum*, triclosan resistance to *Salmonella enterica*, and *pbp4B*, a penicillin-binding protein and the target of  $\beta$ -lactam antibiotics and *marA*, are shared only with the Hadza cohort. *Pbp2*, a point mutation in *N. meningitidis* which confers resistance to  $\beta$ -lactam and a penicillin-binding protein found in *Streptococcus pneumoniae*, also occurs only in the Hadza dataset. *Cat*, which confers resistance to broad-spectrum phenicol antibiotics by antibiotic inactivation, occurs only in the Italian cohort and designates variants of the chloramphenicol acetyltransferase gene in *Enterococcus faecium*, *Lactococcus lactis*, *Listeria monocytogenes*, *Salmonella typhi* and *S. aureus*. *Cat* has also been detected amongst isolated Amerindians [5].

### Conclusions

In this study, we performed a comprehensive analysis of an ancient palaeo-faecal specimen derived from a fifteenth century Middle Iron Age (Neolithic) South African Bantu-speaking hunter-agro-pastoralist. Although representative of the IM composition, metabolic capacity and ARG configuration of the distal (i.e. the colon including the cecum, rectum and anal canal) IM of a single human individual, following particular dietary consumption and excreted at a single point in time, the characterisation of an authenticated ancient African Bantu-speaker IM is an important step towards understanding the ancestral (i.e. pre-colonial African) state of the human IM. Our analyses designate a diet atypical of what is generally expected from a Neolithic (Iron Age) IM [17], instead comprising taxa indicative of a mixed forager-agro-pastoralist diet, supporting the role of dietary habits in shaping human IM composition.

It must be emphasised that the Neolithic of South Africa is different from the 'classic' Near-Eastern Neolithic, as foraging and hunting did play a prominent role in the subsistence configuration of southern African Iron Age communities [40]. It must also be emphasised that the

samples derived from the interior of the BRS specimen (i.e. BRS2, BRS3 and BRS4) 'cluster' with the ancient comparative samples (i.e. those derived from Ötzi) in terms of both taxonomic composition (ANOSIM analysis revealed significant differences between the ancient and modern IM samples) ( $R = 0.8361$ ,  $p = < 0.001$ ) for 731 taxa (Fig. S7) and metabolic capacity (ANOSIM analysis revealed significant differences) ( $R = 0.4840$ ,  $p = < 0.001$ ) (Fig. 4a), and not with the sedimentary controls (SC1 and SC2) or the modern comparative (i.e., Hadza, Malawian and Italian) cohorts. In contrast to contemporary (Hadza, Malawian and Italian) and ancient (BRS and Ötzi) human IM taxonomic composition and metabolic capacity, it is evident that the changes brought about by modern human dietary composition, exposure to toxic pollutants and the excessive use of antibiotics, almost certainly resulted in positive selection for bacterial taxa involved in specific metabolic IM activities [69, 77]. While this does not correlate directly with geography, it does exhibit a temporal trend towards the selection of KO genes in direct response to a number of specific changes in human dietary behaviour and environmental interaction and modification.

The IM of the BRS individual represents a unique taxonomic and metabolic configuration not observed in either contemporary African, or European, populations. Several studies have found that IM composition differs between Western urbanised and indigenous rural populations, and that these dissimilarities frequently correlate with dietary characteristics. In this instance, the diet of the BRS individual, based on hunting, foraging and also agricultural and pastoral resources, differs from the typical Western diet comprising preservatives and food-enhancers, as well as coffee, chocolate, soy, wine and citrus. In terms of modern human hygiene practices, it has been suggested that regular contact with 'old friends' (including both pathogenic and commensal environmental bacteria) is significantly diminished in Western countries [78] and that, given our extensive evolutionary history with microbes [19], this diminishes the capacity of the modern human IM to mediate allergic reactions and autoimmune and inflammatory diseases. It is evident that the ubiquitous use of antibiotics has altered the properties of formerly commensal bacteria, and of the human IM [79]. We therefore hypothesise that, by compelling commensal IM residents to respond to the introduction of antibiotics prescribed for pathogenic taxa, artificially introduced dysbiosis has significantly modulated the pathogenic potential of commensal taxa, resulting in long-term deleterious impacts on optimal human IM functioning.

The IM of the BRS individual also provides evidence for recent human IM adaptation to environmental pollutants. The emergence of xenobiotic degradation

pathways involved in naphthalene, chloroalkane and chloroalkene, benzoate and xylene degradation is likely a population-wide functional response of the IM to exposure to toxic and foreign compounds that are so ubiquitous in industrialised urban environments. The respective enrichment and depletion of several KO genes implicated in the metabolism of morphine and codeine, as well as additives and supplements including pyruvate, L-arginine and beta-alanine, are also indicative of the adaptive capacity of the human IM. Given these modern influences, the contemporary human IM appears to be predisposed towards shifting to a state of dysbiosis. Such altered states of equilibrium frequently result in the pathogenesis of inflammatory diseases and infections, including autoimmune and allergic diseases, obesity and diabetes. The IM of the BRS individual also corroborates the premise that ARGs are a feature of the human IM, regardless of exposure to currently available commercial antibiotics.

The large number of taxonomically (92.63%) and metabolically (88.51%) unassigned reads in the BRS palaeo-faecal specimen analysed here, granting that this might, to some extent, be a result of aDNA damage and the inability to 'map' all reads to existing comparative sequences [44, 45] is suggestive of substantial unknown IM taxonomic diversity and metabolic functionality (Table S12). In the future, the identification of these taxa and their respective metabolic capacities might have significant implications for identifying health risks specific to the sub-Saharan African Bantu-speaker population, which has increased in prevalence with the adoption of Western diets, medical treatments and exposure to modern pollutants. Given that sub-Saharan Africans living outside Africa exhibit a high prevalence of complex diseases, the comparison of ancient African IM data to those of modern Africans might facilitate not only retrospective disease diagnosis, but also the identification of IM-related risk factors that contribute to the onset of certain diseases.

## Methods

### Accelerator mass spectrometry dating

Two sub-samples derived from the interior of the specimen were subjected to accelerator mass spectrometry (AMS) dating. The samples were pre-treated using the standard acid-base-acid approach [80] performed at 70 °C. Carbon was oxidised using off-line combustion in the presence of excess CuO and Ag, and the resulting CO<sub>2</sub> was reduced to graphite through Fe reduction at 600 °C [81]. The graphite was measured at the iThemba LABS AMS facility using oxalic acid II and coal as the reference and background, respectively. We report <sup>14</sup>C ages in conventional radiocarbon years BP (i.e. before present refers to AD 1950). Dates were calibrated using



'Calib 7.1' [82] using the 'SHcal13' dataset [83] and are reported in years AD (anno domini).

### Dietary isotope analyses

To investigate the dietary composition of the BRS individual, one sub-sample derived from the interior of the specimen was subjected to isotopic analyses. The sample was homogenised using a mortar and pestle and then divided into three sub-samples. The first was left untreated, and the second was subjected to a lipid extraction process using a 2:1 chloroform/ethanol solution to remove any lipids present [84]. The sample was covered with 25 ml of the solution, and the mixture agitated in an ultra-sonic bath for 15 min and then left overnight. The solvent was then decanted, and the sample dried at 70 °C prior to weighing for analysis. The third sample was covered with 25 ml 1% hydrochloric acid (HCl) to remove inorganic carbonates [78], agitated for 15 min and left overnight. The acid was then decanted, and the sample repeatedly washed (6 times) with distilled water before drying at 70 °C. Aliquots of the samples weighing between 0.80 mg and 0.90 mg were weighed using a Mettler Toledo MX5 micro-balance. The weighed material was placed in tin capsules that had been pre-cleaned in toluene. All the samples were run in triplicate. Samples for isotopic analyses were combusted at 1020 °C using an elemental analyser (Flash EA 1112 Series) coupled to a Delta V Plus stable light isotope ratio mass spectrometer via a ConFlo IV system (Thermo Fischer, Bremen, Germany), housed at the Stable Isotope Laboratory, University of Pretoria. Two laboratory running standards (Merck Gel:  $\delta^{13}\text{C} = -20.26\text{‰}$ ,  $\delta^{15}\text{N} = 7.89\text{‰}$ , C% = 41.28, N% = 15.29 and DL-Valine:  $\delta^{13}\text{C} = -10.57\text{‰}$ ,  $\delta^{15}\text{N} = -6.15\text{‰}$ , C% = 55.50, N% = 11.86) and a blank sample were run after every 11 unknown samples. Data corrections were performed using the values obtained for the Merck Gel during each run, and the values for the DL-Valine standard provide the  $\pm$  error/precision for each run. The precision for the BRS analyses was  $> 0.04\text{‰}$  and  $0.05\text{‰}$  for nitrogen and carbon, respectively. These running standards are calibrated against international standards, i.e. National Institute of Standards and Technology (NIST): NIST 1557b (bovine liver), NIST 2976 (muscle tissue) and NIST 1547 (peach leaves). All results are referenced to Vienna Pee-Dee Belmontite for carbon isotope values and to air for nitrogen isotope values. Results are expressed in delta notation using a per mille scale using the standard equation ' $\delta X(\text{‰}) = ((R_{\text{sample}}/R_{\text{standard}}) - 1)$ ' where  $X = ^{15}\text{N}$  or  $^{13}\text{C}$  and  $R$  represents  $^{15}\text{N}/^{14}\text{N}$  or  $^{13}\text{C}/^{12}\text{C}$  respectively.

### Intestinal parasite detection

In addition to genomic taxonomic profiling, we also performed microscopic analysis to determine the incidence

of intestinal parasitic helminths and protozoa. The extraction protocols applied in palaeo-parasitology used to extract parasitic markers (i.e. eggs or oocysts) typically entail rehydration, homogenisation and micro-sieving [85]. The sub-sample ( $\sim 5$  g) was placed in a rehydration solution comprising 50 ml 0.5% trisodium phosphate (TSP) solution and 50 ml 5% glycerinated water for 7 days, after which it was ground and passed through an ultrasonic bath for 1 min. The sample was then filtered in a sieving column comprising mesh sizes of 315  $\mu\text{m}$ , 160  $\mu\text{m}$ , 50  $\mu\text{m}$  and 25  $\mu\text{m}$  in aperture diameter. Because of the typical size of most intestinal parasite eggs range from 30 to 160  $\mu\text{m}$  long and 15 to 90  $\mu\text{m}$  wide, only the two last sieves (i.e. 50  $\mu\text{m}$  and 25  $\mu\text{m}$ ) were subjected to microscopic analyses.

### Scanning electron microscopy

We immobilised 0.5 g palaeo-faecal material on double-sided carbon tape (SPI supplies). Excess loose particles were blown off with compressed argon gas and coated with gold using an Emitech K450X sputter coater (Quorum Technologies, UK). Scanning electron microscopy (SEM) images were acquired on a Zeiss Ultra Plus Field Emission Scanning Electron microscopes (Carl Zeiss, Oberkochen, Germany), at an accelerating voltage of 1 kV.

### Ancient DNA extraction and library preparation

All pre-PCR amplification steps were carried out in dedicated aDNA 'clean' laboratories at the Centre for GeoGenetics in Copenhagen, Denmark, applying strict aDNA practices and established aDNA protocols, as per Seersholm et al. [86] and Warinner et al. [87]. Extractions for shotgun metagenome sequencing were carried out using a phenol-chloroform- and kit-based extraction protocol optimised, by the Centre for GeoGenetics, for ancient sedimentary and faecal samples. This entailed dissolving  $\sim 16$  g of palaeo-faecal matter in 40-ml digestion buffer (18 mM EDTA, 100  $\mu\text{g ml}^{-1}$  proteinase K, 7% *N*-lauryl-sarcosyl, 50 mM dithiothreitol and 3% mercaptoethanol) homogenised for  $2 \times 20$  s on a FastPrep-24 (MP Biomedicals) and incubated overnight at 37 °C. Following incubation, samples were spun down, and the supernatant transferred to a clean 15-ml tube. Next, samples were subjected to an inhibitor removal step using buffers C2 and C3 from PowerMax Soil DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA) followed by DNA isolation using phenol and chloroform in a ratio of 1:2. Finally, extracts were concentrated and purified with a 15-ml 30-kDa Amicon Ultra-4 Centrifugal Filter (Millipore) using two wash steps with 750  $\mu\text{l}$  buffer C6 from PowerMax Soil DNA Isolation Kit. Extracts were stored at  $-20$  °C in 50–100  $\mu\text{l}$  C6 buffer. Seven libraries, comprising BRS1



(SC1), BRS2, BRS3, BRS4, BRS5 (SC2) and two negative controls (i.e. an extraction and library preparation control referred to as 'E-LPCs'), were constructed using the NEBNext DNA Library Prep Master Mix for 454 (E6070), with the following modifications: 1–20 µl extract was used for the end-repair step depending on the DNA concentration and inhibition level. End-repair reactions were incubated for 20 min at 12 °C and 15 min at 37 °C. At the adapter ligation step, reactions were incubated at 20 °C for 15 min. The final fill-in reaction was performed at 60 °C for 20 min followed by 80 °C for another 20 min to inactivate the enzyme. Each reaction step was followed by a purification step with MinElute columns (Qiagen) using 400 µl buffer PB for each 25 µl reaction. Finally, libraries were pooled in equimolar concentrations and sequenced on an Illumina HiSeq 2500 platform at the Danish National High-Throughput DNA Sequencing Centre. The libraries were sequenced twice to improve DNA recovery, producing 14.563 Gbp of data (17,979,669 reads) for BRS1, 29.201 Gbp (36,050,926 reads) for BRS2, 0.709 Gbp (8,754,692 reads) for BRS3, 95.323 Gbp (117,683,541 reads) for BRS4 and 9.54 Gbp (110,485,402 reads) for BRS5. The merged E-LPCs yielded 2.946 Gbp of data (3,637,328 reads) (Table S13).

### Sequence processing and microbial taxonomic profiling

The potential for retrieving ancient IM data from palaeofaecal remains is confounded by technical and biological variables [14, 44, 45]. In technical terms, the detection of > 90 microbial genera in DNA extraction and library preparation controls suggests that reagent and laboratory contamination can influence sequence-based IM analyses [88–90]. The choice of DNA extraction protocols can also impact metagenomic compositional profiles [90, 91]. In biological terms, IM research generally focuses on the microbial community of the large intestine as expressed in stools, despite the fact that the 6.5 m human digestive tract consists of three organs, i.e. the stomach, small intestine and large intestine [6]. Since microbial communities change along the length of the GI, differences exist between oral, intestinal and faecal taxonomic profiles in both modern [91, 92] and ancient [25] instances. Moreover, and besides the influence on IM taxonomic composition of diet [57], age [61], seasonal variation [92] and host immuno-modulation [93], stool consistency also influence IM taxonomic composition [94]. Differences in taxonomic composition between the cores and cortices of specimens have also been documented, with larger proportions of soil-derived taxa present in the cortices [24]. In post-depositional terms, the retrieval of ancient IM data is confounded by on-going microbial activity and also environmental contamination [95–97]. Instances of reverse contamination, i.e. from a faecal specimen into the

surrounding sediment, are also probable as the exchange of microbes between a stool, and the surrounding sediment would certainly occur.

With these concerns in mind, taxonomic profiling was preceded by several data pre-processing steps. First, raw sequence reads were processed to remove all Illumina PhiX spikes, human reads and all exact duplicate reads present in the extraction ( $n = 1$ ) and library preparation ( $n = 1$ ) negative controls (E-LPCs) using BBDuk [98]. Second, barcodes, adapters, reads shorter than 25 base-pairs [bp], and 'quality score' < 25 were removed from the dataset [98, 99] using AdapterRemoval V2 [100]. Taxonomic binning was then performed via comparison of the shotgun reads with the BLASTn v2.2.31 + NCBI *nt* database [101, 102]. Taxa were identified using MEGAN CE v6.10.10 [103] by using the weighted lowest common ancestor ('wlca') option and the default percent-to-cover value setting ('80') with parameter values set as follows: min bit score 50, expect value (*e* value)  $1.0e-10$ , top percent 10, min support 10 and min complexity 0.45. Species identifications were based on significant hits (bit score  $\geq 50$ ) and on MEGAN parameters established at 'identities' 97%, 'positives' 100% and no (0%) 'gaps'. Comparisons of BRS IM sequence reads with those derived from other (comparative) IMs were performed by the sub-sampling in of the reads to the lowest number of reads present in any sample (i.e. BRS1 (SC1) with 56,682 filtered sequence reads).

### Ancient DNA authentication

Molecular damage following death is a standard feature of all aDNA molecules. The accumulation of deaminated cytosine [uracil] within the overhanging ends of aDNA templates results in increasing cytosine (C) to thymine (T) misincorporation rates toward read starts, with matching guanine (G) to adenine (A) misincorporations increasing toward read ends in double-stranded library preparations [104, 105]. MapDamage [38, 39] is widely used to determine the incidence of cytosine (C) to thymine (T) and guanine (G) to adenine (A) substitution rates at the 5'-ends and 3'-ends of strands. MapDamage is not however optimised for ancient samples lacking high genome coverage that would permit the identification of all possible misincorporations [105], and DNA damage patterns cannot be calculated for taxa with insufficient read counts, i.e. < 150 [105, 106]. In addition, DNA fragmentation rates vary according to environmental conditions and the types of organisms involved [104–106], often resulting in 'alternative' damage patterns [103–108]. We therefore also validated the antiquity of putatively ancient taxa by a statistical method that compares post-mortem damage patterns indicative of the cytosine to thymine (C-T) substitutions at the 5'-ends of sequence reads [37, 109]. High-quality filtered reads

were aligned to comparative genomes (Table S13) using BWA ( $-n$  0.02,  $-l$  1024) [109], and duplicate sequence reads were removed using the Picard tools script Mark-Duplicates (<https://broadinstitute.github.io/picard/>). Resulting alignments were used to perform statistical DNA damage estimation analyses which entailed the calculation of goodness-of-fit  $p$  values ( $p < 0.05$ ) indicative of significant cytosine to thymine (C-T) substitutions at the 5'-ends of sequence reads using PMDtools (<https://omictools.com/pmdtools-tool>) [37].

The authentication of aDNA sequence reads was furthermore based on the comparison of reads derived from negative (i.e. extraction and library preparation) controls [87, 96, 97, 103, 104]. In addition to the detection of > 90 microbial taxa derived from reagents and laboratory contamination [88, 89], the probability that negative controls can become cross-contaminated during sample processing [105] complicates the authentication process. To establish the antiquity of microbial taxa occurring in the E-LPCs (i.e. *Arthrobacter*, *Blautia*, *Klebsiella*, *Lactobacillus*, *Prevotella* and *Ruminococcus*), we compared the read yields in the BRS specimen with those in the E-LPCs. Since aDNA sequences are shorter than those derived from modern organisms, most frequently via contamination [88, 89, 107, 108, 110–112], DNA read length was also used as criteria in the authentication process. This criterion is particularly relevant in instances that exclude the use of bleach for the removal of contaminants [86, 87], as is the case here. Lastly, evaluating ecological conformity, i.e. excluding DNA reads that are derived from either non-indigenous taxa, foreign contaminants or false-positive identifications, such as *Apteryx* (kiwi), *Cyprinus* (carp), *Oncorhynchus* (salmon) and *Oryza* (rice), was used to assess taxonomic community composition for biological plausibility [96, 97]. The authenticity (i.e. prehistoric provenience) of microbial and macrobial sequence-derived taxa was therefore evaluated according to (1) validating the existence of C-T and G-A substitution rates, (2) statistical aDNA sequence damage estimation, (3) comparison to negative DNA E-LPCs, (4) DNA read length characteristics and (5) ecological conformity.

### Functional metabolic profiling

To discern the metabolic pathways associated with the dietary and environmental factors characteristic of the Bantu-speaking forager-agro-pastoralist in question, we identified reads assigned to functional genes in the shotgun metagenome dataset. To ascertain the presence of microbial groups either positively or negatively correlated with specific metabolic profiles, we determined functional categories based on DIAMOND v0.36.98 BLASTx comparisons to the NCBI non-redundant protein database. The GI accessions were used to identify

the Kyoto Encyclopaedia of Genes and Genomes (KEGG) orthologies (<https://www.genome.jp/kegg/>) in MEGAN CE v6.10.10 [103] by using the weighted lowest common ancestor ['wlca'] option and the default percent-to-cover value setting ('80') with the parameters as follows: 'e value cut-off'  $1.0e-4$ , 'top percent' 10, 'min support' 10, 'min complexity' 0.45 and 'identity' 97%. Comparisons of KEGG orthologies were performed following the sub-sampling to the lowest number of reads for any sample (i.e. BRS1 (SC1) with 56,682 sequence reads).

### Comparative IM datasets

To gain insight into the taxonomic and functional (metabolic) differences between the BRS IM ( $n = 4$ ) and other ancient and modern (including 'traditional') IMs, we compared BRS with data derived from the shotgun metagenome sequencing of Malawian agro-pastoralists [51] (MG-RAST <http://metagenomics.anl.gov/> accession number 'qiime:621'), Tanzanian Hadza hunter-gatherers [26] (NCBI SRA SRP056480 Bioproject ID PRJNA278393), contemporary Italians [26] and the Copper Age (dated to c. 3250 BC) Alpine (Tyrolean) Iceman (Ötzi) [112] (European Nucleotide Archive accession number ERP012908). For comparison, four metagenome samples (comprising two males and two females) were randomly selected from each of the comparative cohorts ( $n = 12$ ).

### Detecting antibiotic resistance genes

The resistome is defined as the complete set of antibiotic resistant genes (ARGs) presented in a microbial community, which is important for understanding the proliferation of pathogen antibiotic resistance [26]. Sequence reads were aligned to the MEGARes database [113] using the Burrows-Wheeler Alignment tool (BWA) [109]. The BRS IM resistome was analysed using Resistome Analyser (<https://github.com/cdeanj/resistomeanalyzer>) by applying the default threshold of '80' to determine the gene significance and in order to decrease false positive gene identifications. Relative abundance for each of the resistance genes was calculated and sub-sampled to the lowest number of sequences (i.e. the Ötzi 'small intestine' (Ötzi\_SI\_F) comprising 2377 sequences) in any sample (i.e. 28,524 sequences for a total of twelve samples). Because of low numbers of gene assignments, seven samples were excluded from the analysis.

### Statistical analyses

Statistical analysis of the ancient and modern IM samples was performed by filtering the taxa and KEGG orthologies for > 3 occurrences in at least 20% of the samples. The data was Hellinger transformed, and the Bray-Curtis dissimilarity matrix was used for the *vegdist*

and *anosim* functions in the VEGAN (<https://www.rdocumentation.org/packages/vegan/versions/2.4-2>) package of R. The ordination plots were generated in the Phyloseq package [114] of R. Functional (metabolic) group significance tests were performed in Qiime v1.9.1 package [115] for the ancient and modern comparative cohorts, and only gene categories with significant *p* value ( $p < 0.05$ ) were included in the analyses. The *p* values were corrected using false discovery rate (FDR) (*q*) [116], and Kruskal-Wallis values (*H*) were used to determine the significance of differences between samples. Hierarchical clustering using complete-linkage based on Spearman's correlations was performed and visualised in R using 'gplots' (<https://www.rdocumentation.org/packages/gplots/versions/3.0.1>). Heat maps were generated using Spearman's correlation and complete linkage method for microbial taxa and antibiotic resistance genes. Taxa were filtered for occurrence of > 3 in at least 20% of the samples, and ARG data was filtered for the occurrence of > 5 in at least 20% samples. The heat map for the KEGG orthologies linked to the twenty-four ancient IM taxa (1487 categories) was produced using Spearman's correlation and the Ward-linkage method.

## Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s40168-020-00832-x>.

**Additional file 1: Figure S1.** Additional information concerning the archaeological provenance of the BRS faecal specimen. **Figure S2.** Processing of the faecal specimen at the Centre for GeoGenetics, Copenhagen, Denmark. **Figure S3.** Dot-plot indicating the occurrence of statistically-significant C-T *p*-values. **Figure S4.** Biplot of  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  stable isotope values obtained for the BRS specimen. **Figure S5.** SEM analysis detected bacterial cells, plant fragments and saprophytic organisms. **Figure S6.** Heat-map indicating differences in taxonomic community structure for IM datasets. **Figure S7.** Comparing 'relative abundance' and 'presence-absence' as taxonomic representation. **Figure S8.** Comparison of the incidence of the twenty-four authenticated ancient IM taxa. **Figure S9.** Heat-map indicating the presence of fifteen functional ARGs identified.

**Additional file 2: Table S1.** Sequence reads for environmental- and subsistence-related taxa detected. **Table S2.** Information concerning  $^{14}\text{C}$  Accelerator Mass Spectrometry (AMS) dating. **Table S3a.** Processing protocol and results for isotope analyses. **Table S3b.** Results for isotope analyses (Merck standard). **Table S3c.** Results for isotope analyses (DL-Valine standard). **Table S4.** Abundance of bacterial taxonomic categories in the IM datasets. **Table S5.** Sequence read-length distribution for taxa identified in this study. **Table S6.** Significant KEGG pathways in the comparative IM datasets analysed. **Table S7.** Relative abundance of eighteen significant KEGG pathways in the IM cohorts. **Table S8.** Enrichment and depletion of KO metabolic gene categories in the comparative IM sample cohorts based on *p*-value ( $p < 0.05$ ) designation. **Table S9.** Enrichment and depletion of KO metabolic gene categories in the comparative IM sample cohorts based on false discovery rate (FDR) corrected *p*-values ( $q < 0.05$ ). **Table S10.** Enrichment and depletion of KO metabolic gene categories in the ancient and modern comparative IM sample cohort as calculated for the twenty-four authenticated ancient IM taxa. **Table S11.** Comparison of relative abundance of antibiotic resistance genes in the comparative IM cohorts. **Table S12.** Raw and filtered high-quality sequence read counts as related to the comparative IM datasets. **Table**

**S13.** Information concerning the comparative NCBI genomes used during this study.

## Acknowledgements

We thank Clemens Weiß (University of Tübingen), Fredrik Seersholm (University of Copenhagen) and Pedro Lebre and Eudri Venter (University of Pretoria) for informative discussions and analytical support.

## Authors' contributions

RFR, AJH, JBR and SV conceived the study and composed the manuscript. SV performed the bioinformatic and statistical analyses, and RFR and SV generated the figures. JBR, ARI, TBB and RFR performed aDNA extractions, library preparations and sequencing. GP and AV provided access to the analysed specimen. GH performed the isotope analyses, and SW performed the  $^{14}\text{C}$  dating. MLB carried out intestinal parasitic analyses. MP and RFR performed SEM analyses. EW, DAC and YVDP provided funding and access to analytical facilities. AJH, JBR and RFR conceived the analytical protocol, and AJH supervised the experimental protocol. All authors contributed to the completion of the final manuscript. The author(s) read and approved the final manuscript.

## Funding

RFR acknowledges funding provided by a National Geographic Society Scientific Exploration Grant (Nr. NGS-371R-18) and by the Oppenheimer Endowed Fellowship in Molecular Archaeology. GP, AV, RFR and JBR acknowledge funding provided by the French Ministère des Affaires Étrangères and the French Institute of South Africa (IFAS), and RFR and JBR acknowledge research mobility assistance provided by the NRF (South Africa UID Nr. 105197) and DRIS (University of Pretoria). We thank the South African Heritage Resources Agency (SAHRA) for the provision of export permits (Permit ID: 2364).

## Availability of data and materials

Comparative data is derived from the shotgun metagenome sequencing of Malawian agro-pastoralists (MG-RAST <http://metagenomics.anl.gov/> accession number 'qiime:621'), Tanzanian Hadza hunter-gatherers (NCBI SRA SRP056480 Bioproject ID PRJNA278393), contemporary Italians and the Copper Age (dated to c. 3,250 BC) Alpine (Tyrolean) Iceman (Ötzi) (European Nucleotide Archive accession number ERP012908). All data generated during this study are included in this published article (and its supplementary information files). Sequencing data generated in this study can be accessed via the BioProject accession number PRJNA527177 in the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>). The short read archive accession (SRA) numbers for samples BRS1, BRS2, BRS3, BRS4 and BRS5 are SRR8731956, SRR8731955, SRR8731958, SRR8731957 and SRR8731959, respectively.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests. The funding sponsors had no role in the design of the study, the collection, analyses and interpretation of data, in the writing of the manuscript or in the decision to distribute the results.

## Author details

<sup>1</sup>Center for Microbial Ecology and Genomics, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Hatfield, South Africa. <sup>2</sup>Department of Anthropology and Geography, Human Origins and Palaeoenvironmental Research Group, Oxford Brookes University, Oxford, UK. <sup>3</sup>Department of Molecular Genetics and Microbiology, Pontificia Universidad Católica de Chile, Santiago, Chile. <sup>4</sup>Centre for GeoGenetics, GLOBE Institute, University of Copenhagen, Hatfield, Denmark. <sup>5</sup>CNRS, UMR 7041 ArScAn-AnTET, Université Paris-Nanterre, Paris, France. <sup>6</sup>Evolutionary Studies Institute, University of the Witwatersrand, Braamfontein Johannesburg, South Africa. <sup>7</sup>Department of Early Prehistory and Quaternary Ecology, University of Tübingen, Tübingen, Germany. <sup>8</sup>Mammal Research Institute, University of Pretoria, Hatfield, South Africa. <sup>9</sup>Themba LABS, Braamfontein Johannesburg, South Africa. <sup>10</sup>University of Bourgogne France-Comte, CNRS UMR 6249

Chrono-environment, Besancon, France. <sup>11</sup>VIB Centre for Plant Systems Biology, Ghent, Belgium. <sup>12</sup>Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium. <sup>13</sup>GeoGenetics Group, Department of Zoology, University of Cambridge, Cambridge, UK. <sup>14</sup>Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.

Received: 26 November 2019 Accepted: 18 March 2020

Published online: 06 May 2020

## References

- Lederberg JJ, McCray AT. 'Ome sweet' omics: a genealogical treasury of words. *Scientist*. 2001;15:8.
- Warinner C, Speller C, Collins MJ, Lewis CM. Ancient human microbiomes. *J Hum Evol*. 2015;79:125–6.
- Thursby E, Juge N. Introduction to the human gut microbiota. *Biochem J*. 2017;474:1823–6.
- van den Elsen LW, Poyntz HC, Weyrich LS, Young W, Forbes-Blom EE. Embracing the gut microbiota: the new frontier for inflammatory and infectious diseases. *Clin Transl Immunology*. 2017. <https://doi.org/10.1038/cti.2016.91>.
- Clemente JC, Pehrsson EC, Blaser MJ, Sandhu K, Gao Z, Wang B, Magris M, Hidalgo G, Contreras M, Noya-Alarcón O, Lander O, McDonald J, Cox M, Walter J, Oh PL, Ruiz JF, Rodriguez S, Shen N, Song SJ, Metcalf J, Knight R, Dantas G, Dominguez-Bello MG. The microbiome of uncontacted Amerindians. *Sci Adv*. 2015. <https://doi.org/10.1126/sciadv.1500183>.
- Davenport ER, Sanders JG, Song SJ, Amato KR, Clark AG, Knight R. The human microbiome in evolution. *BMC Biol*. 2017. <https://doi.org/10.1186/s12915-017-0454-7>.
- Blaser MJ. The past and future biology of the human microbiome in an age of extinctions. *Cell*. 2018;172:1173–7.
- Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G, Turrioni S, Biagi E, Peano C, Severgnini M, Fiori J, Gotti R, De Bellis G, Luiselli D, Brigidi P, Mabulla A, Marlowe F, Henry AG, Crittenden AN. Gut microbiome of the Hadza hunter-gatherers. *Nat Commun*. 2014. <https://doi.org/10.1038/ncomms4654>.
- Obregon-Tito AJ, Tito RY, Metcalf J, Sankaranarayanan K, Clemente JC, Ursell LK, Zech Xu Z, Van Treuren W, Knight R, Gaffney PM, Spicer P, Lawson P, Marin-Reyes L, Trujillo-Villarroel O, Foster M, Gujja-Poma E, Troncoso-Corzo L, Warinner C, Ozga AT, Lewis CM. Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat Commun*. 2015. <https://doi.org/10.1038/ncomms7505>.
- Warinner C, Lewis CM. Microbiome and health in past and present human populations. *Am Anthropol*. 2015;117:740–1.
- Lee R, Daly RH. Cambridge encyclopaedia of hunters and gatherers. Cambridge: Cambridge University Press; 1999. ISBN 9780521609197.
- Gomez A, Petzelkova KJ, Burns MB, Yeoman CJ, Amato AR, Vlckova K, Modry D, Todd A, Jost Robinson CA, Remis MJ, Torralba MG, Morton E, Umaña JD, Carbonero F, Gaskins HR, Nelson KE, Wilson BA, Stumpf RM, White BA, Leigh SR. Gut microbiome of coexisting BaAka Pygmies and Bantu reflects gradients of traditional subsistence patterns. *Cell Rep*. 2016;14:2142–53.
- Girard C, Tromas N, Amyot M, Shapiro BJ. Gut microbiome of the Canadian Arctic Inuit. *mSphere*. 2017. <https://doi.org/10.1128/mSphere.00297-16>.
- Voigt AY, Costea PI, Kultima JR, Li SS, Zeller G, Sunagawa S, Bork P. Temporal and technical variability of human gut metagenomes. *Genome Biol*. 2015. <https://doi.org/10.1186/s13059-015-0639-8>.
- Walter J, Ley R. The human gut microbiome: ecology and recent evolutionary changes. *Annu Rev Microbiol*. 2011;65:411–29.
- Blaser MJ, Falkow S. What are the consequences of the disappearing human microbiota? *Nat Rev Microbiol*. 2009;7:887–94.
- Adler CJ, Dobney K, Weyrich LS, Kaidonis J, Walker AW, Haak W, Bradshaw CJ, Townsend G, Soltysiak A, Alt KW, Parkhill J, Cooper A. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat Genet*. 2013;45:450–5.
- Quercia S, Candela M, Giuliani C, Turrioni S, Luiselli D, Rampelli S, Brigidi P, Franceschi C, Bacalini MG, Garagnani P, Pirazzini C. From lifetime to evolution: timescales of human gut microbiota adaptation. *Front Microbiol*. 2014. <https://doi.org/10.3389/fmicb.2014.00587>.
- Moeller AH, Caro-Quintero A, Mjunga D, Georgiev AV, Lonsdorf EV, Muller MN, Pusey AE, Peeters M, Hahn BH, Ochman H. Cospeciation of gut microbiota with hominids. *Science*. 2016;353:380–2.
- Reyes-Centeno H, Harvati K, Jäger G. Tracking modern human population history from linguistic and cranial phenotype. *Sci Rep*. 2016. <https://doi.org/10.1038/srep36645>.
- Houldcroft C, Ramond JB, Rifkin RF, Underdown SJ. Migrating microbes: what pathogens can tell us about population movements and human evolution. *Ann Hum Biol*. 2017;44:397–407.
- Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, Falush D, Stamer C, Prugnolle F, van der Merwe SW, Yamaoka Y, Graham DY, Perez-Trallero E, Wadstrom T, Suerbaum S, Achtman M. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature*. 2007;445:915–8.
- Gilbert MT, Jenkins DL, Götherström A, Naveran N, Sanchez JJ, Hofreiter M, Thomsen PF, Binladen J, Higham TF, Yohe RM, Parr R, Cummings LS, Willerslev E. DNA from pre-Clovis human coprolites in Oregon, North America. *Science*. 2008;320:786–9.
- Cano RJ, Rivera-Perez J, Toranzos GA, Santiago-Rodriguez TM, Narganes-Storde YM, Chanlatte-Baik L, García-Roldán E, Bunkley-Williams L, Massey SE. Paleomicrobiology: revealing fecal microbiomes of ancient indigenous cultures. *PLoS One*. 2014. <https://doi.org/10.1371/journal.pone.0106833>.
- Santiago-Rodriguez TM, Fornaciari G, Luciani S, Dowd SE, Toranzos GA, Marota I, Cano RJ. Gut microbiome of an 11<sup>th</sup> century A.D. Pre-Columbian Andean mummy. *PLoS One*. 2015. <https://doi.org/10.1371/journal.pone.0138135>.
- Rampelli S, Schnorr SL, Consolandi C, Turrioni S, Severgnini M, Peano C, Brigidi P, Crittenden AN, Henry AG, Candela M. Metagenome sequencing of the Hadza hunter-gatherer gut microbiota. *Curr Biol*. 2015;25:1682–93.
- Porraz G, Val A, Dayet L, de la Peña P, Douze K, Miller CE, Murungi M, Tribolo C, Schmid V, Sievers C. Bushman Rock Shelter (Limpopo, South Africa): a perspective from the edge of the Highveld. *S Afr Archaeol Bull*. 2015;70:166–79.
- Russell T, Silva F, Steele J. Modelling the spread of farming in the Bantu-speaking regions of Africa: an archaeology-based phylogeography. *PLoS One*. 2014. <https://doi.org/10.1371/journal.pone.0087854>.
- Porraz G, Val A, Tribolo C, Mercier N, de la Peña P, Haaland MM, Igreja M, Miller CE, Schmid V. The MISS Pietersburg at '28' Bushman Rock Shelter, Limpopo Province, South Africa. *PLoS One*. 2018. <https://doi.org/10.1371/journal.pone.0202853>.
- Wood JR, Wilmshurst JM. A protocol for subsampling Late Quaternary coprolites for multi-proxy analysis. *Quat Sci Rev*. 2014;138:1–5.
- Walker C. Signs of the wild: a field guide to the spoor and signs of the mammals of southern Africa. Struik: Cape Town; 2015.
- Estes RD. The behavior guide to African mammals: including hoofed mammals, carnivores, primates. Berkeley: University of California Press; 2012.
- Huffman TN. Mapungubwe and Great Zimbabwe: the origin and spread of social complexity in southern Africa. *J Anthropol Archaeol*. 2007;28:37–54.
- Samuelson M. Rendering the Cape-as-port: sea-mountain, Cape of Storms/Good Hope, Adamastor and local-world literary formations. *J S Afr Stud*. 2016;42:523–37.
- Campana MG, Robles García N, Rühli FJ, Tuross N. False positives complicate ancient pathogen identifications using high-throughput shotgun sequencing. *BMC Res Notes*. 2014. <https://doi.org/10.1186/1756-0500-7-111>.
- Willerslev E, Davison L, Moora M, Zobel M, Coissac E, Edwards ME, Lorenzen ED, Vestergård M, Gussarova G, Haile J, Craine J, Gielly L, Boessenkool S, Epp LS, Pearman PB, Cheddadi R, Murray D, Bräthen KA, Yoccoz N, Binney H, Cruaud C, Wincker P, Goslar T, Greve Alsos I, Bellemain E, Krag Brysting A, Elven R, Sønstebo JH, Murton J, Sher A, Rasmussen M, Rønn R, Mourier T, Cooper A, Austin J, Möller P, Froese D, Zazula G, Pompanon F, Rioux D, Niderkorn V, Tikhonov A, Savvinov G, Roberts RG, RDE MP, MTP G, Kjær KH, Orlando L, Brochmann C, Taberlet P. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*. 2014;506:47–51.
- Weiß CJ, Dannemann M, Prüfer K, Burbano HA. Contesting the presence of wheat in the British Isles 8,000 years ago by assessing ancient DNA authenticity from low-coverage data. *Elife*. 2015. <https://doi.org/10.7554/eLife.10005>.
- Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics*. 2011;27:2153–5.
- Jónsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L. mapDamage 2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*. 2013;29:1682–4.



40. Badenhorst S. Intensive hunting during the Iron Age of Southern Africa. *J Hum Palaeoecol.* 2015;20:41–5.
41. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. *Genome Med.* 2016;doi:<https://doi.org/10.1186/s13073-016-0307-y>.
42. Qin J, Li R, Raes J, Arumugam M, Solvsten Burgdorf K, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Doré J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, MetaHIT Consortium WJ, Bork P, Ehrlich SD, Wang J. A human gut microbial gene catalog established by metagenomic sequencing. *Nature.* 2010;464:59–65.
43. Derrien M, van Hylckama Vlieg JE. Fate, activity, and impact of ingested bacteria within the human gut microbiota. *Trends Microbiol.* 2015;23:354–66.
44. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. *Front Microbiol.* 2017. <https://doi.org/10.3389/fmicb.2017.02224>.
45. Kistler L, Ware R, Smith O, Collins M, Allaby RG. A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic Acids Res.* 2017;45:6310–20.
46. Velsko IM, Frantz LAF, Herbig A, Larson G, Warinner C. Selection of appropriate metagenome taxonomic classifiers for ancient microbiome research. *mSystems.* 2018. <https://doi.org/10.1128/mSystems.00080-18>.
47. Woolhouse MEJ. Where do emerging pathogens come from? *Microbe.* 2006. <https://doi.org/10.1128/microbe.1.511.1>.
48. Nédélec Y, Sanz J, Baharian SZA, Pacis A, Dumaine A, Grenier JC, Freiman A, Sams AJ, Hebert S, Pagé Sabourin A, Luca F, Blekhan R, Hernandez RD, Pique-Regi R, Tung J, Yotova V, Barreiro LB. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell.* 2017;167:657–69.
49. Kessler SE, Bonnell TR, Byrne RW, Chapman CA. Selection to outsmart the germs: the evolution of disease recognition and social cognition. *J Hum Evol.* 2017;108:92–109.
50. Thornhill R, Fincher CL. The parasite-stress theory of values and sociality. London: Springer; 2014.
51. Pierce ES. Could *Mycobacterium avium* subspecies *paratuberculosis* cause Crohn's disease, ulcerative colitis...and colorectal cancer? *BMC Infect. Agents Cancer.* 2018. <https://doi.org/10.1186/s13027-017-0172-3>.
52. Shin NR, Whon TW, Bae JW. *Proteobacteria*: microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol.* 2015;33:496–503.
53. Hughes ER, Winter MG, Buerkop BA, Spiga L, Furtado de Carvalho T, Zhu W, Gillis CC, Büttner L, Smoot MP, Behrendt CL, Cherry S, Santos RL, Hooper LV, Winter SE. Microbial respiration and formate oxidation as metabolic signatures of inflammation-associated dysbiosis. *Cell Host Microbe.* 2017;21:208–19.
54. Ley RE, Turnbaugh PJ, Klein S, Gordon JL. Microbial ecology: human gut microbes associated with obesity. *Nature.* 2006;444:1022–3.
55. Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on human health: an integrative view. *Cell.* 2012;148:1258–70.
56. Marchesi JR, Adams DH, Fava F, Hermes GD, Hirschfield GM, Hold G, Quraishi MN, Kinross L, Smidt H, Tuohy KM, Thomas LV, Zoetendal EG, Hart A. The gut microbiota and host health: a new clinical frontier. *Gut.* 2016;65:330–9.
57. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poulet JB, Massart S, Collini S, Pieraccini G, Lionetti P. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A.* 2010;107:14691–6.
58. Maier TV, Lucio M, Lee LH, VerBerkmoes NC, Brislawn CJ, Bernhardt J, Lamendella R, McDermott JE, Bergeron N, Heinzmann SS, Morton JT, González A, Ackermann G, Knight R, Riedel K, Krauss RM, Schmitt-Kopplin P, Jansson JK. Impact of dietary resistant starch on the human gut microbiome, metaproteome, and metabolome. *MBio.* 2017. <https://doi.org/10.1128/mBio.01343-17>.
59. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Doré J, Consortium MHIT, Weissenbach J, Ehrlich SD, Bork P. Enterotypes of the human gut microbiome. *Nature.* 2011;473:174–80.
60. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman FD, Lewis JD. Linking long-term dietary patterns with gut microbial enterotypes. *Science.* 2011;334:105–8.
61. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JL. Human gut microbiome viewed across age and geography. *Nature.* 2012;486:222–7.
62. Rodríguez-Vaquero MJ, Alberto MR, Manca de Nadra MC. Antibacterial effect of phenolic compounds from different wines. *Food Control.* 2007;18:93–101.
63. McKenney EA, Ashwell M, Lambert JE, Fellner V. Fecal microbial diversity and putative function in captive western lowland gorillas [*Gorilla gorilla gorilla*], common chimpanzees [*Pan troglodytes*], Hamadryas baboons [*Papio hamadryas*] and binturongs [*Arctictis binturong*]. *Integr Zool.* 2014;9:557–69.
64. Samuel BS, Hansen EE, Manchester JK, Coutinho PM, Henrissat B, Fulton R, Latreille P, Kim K, Wilson RK, Gordon JL. Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proc Natl Acad Sci U S A.* 2007;104:10643–8.
65. Chen T, Long W, Zhang C, Liu S, Zhao L, Hamaker BR. Fiber-utilizing capacity varies in *Prevotella*- versus *Bacteroides*-dominated gut microbiota. *Sci. Rep.* 2017. <https://doi.org/10.1038/s41598-017-02995-4>.
66. Lopetuso LR, Scaldaferrri F, Petito V, Gasbarrini A. Commensal *Clostridia*: leading players in the maintenance of gut homeostasis. *Gut Pathog.* 2013. <https://doi.org/10.1186/1757-4749-5-23>.
67. Chassard C, Delmas E, Robert C, Bernalier-Donadille A. The cellulose-degrading microbial community of the human gut varies according to the presence or absence of methanogens. *FEMS Microbiol Ecol.* 2010;74:205–13.
68. Morton ER, Lynch J, Froment A, Lafosse S, Heyer E, Przeworski M, Blekhan R, Ségurel L. Variation in rural African gut microbiota is strongly correlated with colonization by *Entamoeba* and subsistence. *PLoS Genet.* 2015. <https://doi.org/10.1371/journal.pgen.1005658>.
69. Lozupone CA, Stombaugh JI, Gordon JL, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature.* 2012;489:220–30.
70. Vangay P, Johnson AJ, Ward TL, Al-Ghalith GA, Shields-Cutler RR, Hillmann BM, Lucas SK, Beura LK, Thompson EA, Till LM, Batres R, Paw B, Pergament SL, Saenyakul P, Xiong M, Kim AD, Kim G, Masopust D, Knight D. US immigration westernizes the human gut microbiome. *Cell.* 2018. <https://doi.org/10.1016/j.cell.2018.10.029>.
71. Tomita K, Nagura T, Okuhara Y, Nakajima-Adachi H, Shigematsu N, Aritsuka T, Kaminogawa S, Hachimura S. Dietary melibiose regulates *th* cell response and enhances the induction of oral tolerance. *Biosci Biotechnol Biochem.* 2007;71:2774–80.
72. Fiedler H. Short-chain chlorinated paraffins: production, use and international regulations. In: Boer J, editor. Chlorinated paraffins. The handbook of environmental chemistry. Berlin: Springer; 2010. p. 1–41.
73. Glüge J, Wang Z, Bogdal C, Scheringer M, Hungerbühler K. Global production, use, and emission volumes of short-chain chlorinated paraffins: a minimum scenario. *Sci Total Environ.* 2016;573:1132–46.
74. Paoli A, Bianco A, Grimaldi KA, Lodi A, Bosco G. Long term successful weight loss with a combination biphasic ketogenic Mediterranean diet and Mediterranean diet maintenance protocol. *Nutrients.* 2013;5:5205–17.
75. D'Costa VM, King CE, Kalan L, Morar M, Sung WW, Schwarz C, Froese D, Zazula G, Calmels F, Debruyne R, Golding GB, Poinar HN, Wright GD. Antibiotic resistance is ancient. *Nature.* 2011;477:457–61.
76. Kaur R, Gautam V, Ray P, Singh G, Singhal R, Tiwari R. Daptomycin susceptibility of methicillin resistant *Staphylococcus aureus* (MRSA). *Indian J Med Res.* 2012;136:676–7.
77. Ye K, Gao F, Wang D, Bar-Yosef O, Keinan A. Dietary adaptation of FADS genes in Europe varied across time and geography. *Nat Ecol Evol.* 2017. <https://doi.org/10.1038/s41559-017-0167>.
78. Rook GAW, Brunet LR. Microbes, immunoregulation, and the gut. *Gut.* 2005;54:317–20.
79. Stecher B, Maier L, Hardt WD. 'Blooming' in the gut: how dysbiosis might contribute to pathogen evolution. *Nat Rev Microbiol.* 2013;11:277–84.
80. Stuiver M, Quay PD. Atmospheric  $^{14}\text{C}$  changes resulting from fossil fuel  $\text{CO}_2$  release and cosmic ray flux variability. *Earth and Planetary Science.* 1981;53:349–62.



81. Nemecek M, Wacker L, Gaggeler H. Optimization of the graphitization process at AGE-1. *Radiocarbon*. 2010;52:1380–93.
82. Stuiver M, Reimer PJ, Reimer RW. CALIB 7.1. <http://calib.org> 2020; accessed 7 January 2020.
83. Hogg AG, Hua Q, Blackwell PG, Niu M, Buck CE, Guilderson TP, Heaton TJ, Palmer JG, Reimer PJ, Reimer RW, Turney CSM, Zimmerman SRH. SHCal13 Southern Hemisphere calibration, 0–50,000 years cal BP. *Radiocarbon*. 2013; 55:1889–903.
84. Woodborne S, Huchzermeyer D, Govender D, Pienaar DJ, Hall G, Myburgh JG, Deacon AR, Venter J, Lubcker N. Ecosystem change and the Olifants River crocodile mass mortality events. *Ecosphere*. 2012. <https://doi.org/10.1890/ES12-00170.1>.
85. Dufour B, Le Bailly M. Testing new parasite egg extraction methods in paleoparasitology and an attempt at quantification. *Int J Paleopathol*. 2013; 3:199–203.
86. Seersholm FW, Pedersen MW, Sørensen MJ, Shokry H, Mak SS, Ruter A, Raghavan M, Fitzhugh W, Kjær KH, Willerslev E, Meldgaard M, Kapel CMO, Hansen AJ. DNA evidence of bowhead whale exploitation by Greenlandic Paleo-Inuit 4,000 years ago. *Nat Commun*. 2016. <https://doi.org/10.1038/ncomms13389>.
87. Warinner C, Herbig A, Mann A, Fellows Yates JA, Weiß CL, Burbano HA, Orlando L, Krause J. A robust framework for microbial archaeology. *Annu Rev Genomics Hum Genet*. 2017;18:321–56.
88. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol*. 2014. <https://doi.org/10.1186/s12915-014-0087-z>.
89. Lauder AP, Roche AM, Sherrill-Mix S, Bailey A, Laughlin AL, Bittinger K, Leite R, Elovitz MA, Parry S, Bushman FD. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome*. 2016. <https://doi.org/10.1186/s40168-016-0172-3>.
90. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, Tramontano M, Driessen M, Herczeg R, Jung F, Roat Kultima J, Hayward MR, Coelho LP, Allen-Vercos E, Bertrand L, Blaut M, JRM B, Carton T, Cools-Portier S, Daigneault M, Derrien M, Druetne A, de Vos WM, Finlay BB, Flint HJ, Guarner F, Hattori M, Heilig H, Luna RA, van Hylckama Vlieg J, Junick J, Klymiuk I, Langella P, Le Chatelier E, Mai V, Manichanh C, Martin JC, Mery C, Morita H, O'Toole PW, Orvain C, Raosaheb Patil K, Penders J, Persson S, Pons N, Popova M, Salonen A, Saulnier D, Scott KP, Singh B, Slezak K, Veiga P, Versalovic J, Zhao L, Zoetendal EG, Ehrlich SD, Dore J, Bork P. Towards standards for human fecal sample processing in metagenomic studies. *Nat Biotechnol*. 2017;35:1069–76.
91. Stearns JC, Lynch MDJ, Senadheera DB, Tenenbaum HG, Goldberg MB, Cvitkovitch DG, Croitoru K, Moreno-Hagelsieb G, Neufeld JD. Bacterial biogeography of the human digestive tract. *Sci. Rep*. 2011. <https://doi.org/10.1038/srep00170>.
92. Smits SA, Gonzalez CG, Leach J, Manjurano A, Sonnenburg ED, Lichtman JS, Chantalucha J, Dominguez-Bello MG, Reid G, Knight R, Elias JE, Sonnenburg JL. Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science*. 2017;355:802–6.
93. Donaldson GP, Lee SM, Mazmanian SK. Gut biogeography of the bacterial microbiota. *Nat Rev Microbiol*. 2016;14:20–32.
94. Vandeputte D, Falony G, Vieira-Silva S, Tito RY, Joossens M, Raes J. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *BMJ Gut*. 2016;65: 57–62.
95. Ott SJ, Musfeldt M, Timmis KN, Hampe J, Wenderoth DF, Schreiber S. *In vitro* alterations of intestinal bacterial microbiota in fecal samples during storage. *Diagn Microbiol Infect Dis*. 2004;50:237–45.
96. Tito RY, Macmil S, Wiley G, Najar F, Cleeland L, Qu C, Wang P, Romagne F, Leonard S, Jiménez A, Ruiz A, Reinhard K, Roe BA, Lewis SM Jr. Phylotyping and functional analysis of two ancient human microbiomes. *PLoS One*. 2008. <https://doi.org/10.1371/journal.pone.0003703>.
97. Tito RY, Knights D, Metcalf J, Obregon-Tito AJ, Cleeland L, Najar F, Roe B, Reinhard K, Sobolik K, Belknap S, Foster M, Spicer P, Knight R, Lewis CM Jr. Insights from characterizing extinct human gut microbiomes. *PLoS One*. 2012. <https://doi.org/10.1371/journal.pone.0051146>.
98. Bushnell B. BBMap. 2014; <http://sourceforge.net/projects/bbmap/>.
99. Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One*. 2012. <https://doi.org/10.1371/journal.pone.0034131>.
100. Dabney J, Knapp M, Glocke I, Gansauge M, Weihmann A, Nickel B, Valdiosera C, García N, Pääbo S, Arsuag J, Meyer M. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A*. 2013;110:15758–63.
101. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes*. 2016. <https://doi.org/10.1186/s13104-016-1900-2>.
102. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
103. Huson DH, Beier S, Flade I, Górski A, El-Hadidi M, Mitra S, Ruscheweyh H, Tappu R. MEGAN Community Edition: interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput Biol*. 2016. <https://doi.org/10.1371/journal.pcbi.1004957>.
104. Briggs AW, Stenzel U, Johnson PL, Green RE, Kelso J, Prüfer K, Meyer M, Krause J, Ronan MT, Lachmann M, Pääbo S. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A*. 2007;104: 14616–21.
105. Vågane ÅJ, Herbig A, Campana MG, Robles García NM, Warinner C, Sabin S, Spyrou MA, Andrades Valtueña A, Huson D, Tuross N, Bos KI, Krause J. *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat Ecol Evol*. 2017;2:520–8.
106. Tackney JC, Potter BA, Raff J, Powers M, Watkins WS, Warner D, Reuther JD, Irish JD, O'Rourke DH. Two contemporaneous mitogenomes from terminal Pleistocene burials in eastern Beringia. *Proc Natl Acad Sci U S A*. 2015. <https://doi.org/10.1073/pnas.1511903112>.
107. Dabney J, Meyer M, Pääbo S. Ancient DNA damage. *Mol Ecol Res*. 2017. <https://doi.org/10.1111/1755-0998.12595>.
108. Weiß CL, Schuenemann VJ, Devos J, Shirekar G, Reiter E, Gould BA, Stinchcombe JR, Krause J, Burbano HA. Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. *R Soc Open Sci*. 2016. <https://doi.org/10.1098/rsos.160239>.
109. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
110. Malmström H, Svensson EM, Gilbert MTP, Willerslev E, Götherström A, Holmlund G. More on contamination: the use of asymmetric molecular behavior to identify authentic ancient human DNA. *Mol Biol Evol*. 2007;24: 998–1004.
111. Prüfer K, Stenzel U, Hofreiter M, Pääbo S, Kelso J, Green RE. Computational challenges in the analysis of ancient DNA. *Genome Biol*. 2010. <https://doi.org/10.1186/gb-2010-11-5-r47>.
112. Lugini GA, Milani C, Mancabelli L, Turroni F, Ferrario C, Duranti S, van Sinderen D, Ventura M. Ancient bacteria of the Ötzi's microbiome: a genomic tale from the Copper Age. *Microbiome*. 2017. <https://doi.org/10.1186/s40168-016-0221-y>.
113. Lakin SM, Dean C, Noyes NR, Dettenwanger A, Ross AS, Doster E, Rovira P, Abdo Z, Jones KL, Ruiz J, Belk KL, Morley PS, Boucher C. MEGARes: an antimicrobial resistance database for high throughput sequencing. *Nucleic Acids Res*. 2017. <https://doi.org/10.1093/nar/gkw1009>.
114. McMurdie PJ, Holmes S. Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013. <https://doi.org/10.1371/journal.pone.0061217>.
115. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010. <https://doi.org/10.1038/nmeth.f.303>.
116. Jiang L, Amir A, Morton JT, Heller R, Arias-Castro E, Knight R. Discrete false-discovery rate improves identification of differentially abundant microbes. *mSystems*. 2017. <https://doi.org/10.1128/mSystems.00092-17>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.